Title: Chronic *Salmonella* Infection Contributes to Gallbladder Carcinogenesis

Authors: *Agila Kumari Pragasam[1#], Sonalika Maurya[2#], Kajal Jain[3#], Sujoy Pal[4], Christu Raja[4], Shakti Kumar[1], Ayushi Purohit[1], Dibyabhaba Pradhan[4], Kirti Kajal[2], Daizee Talukdar[1], Anand Narayan Singh[4], Jyoti Verma[1], Pradipta Jana[1], Shaifali Rawat[5], Pallavi Kshetrapal[6], Asuri Krishna[3], Subodh Kumar[7], Virinder Kumar Bansal[3], Rajni Yadav, Bhabatosh Das[1*8], Chittur V Srikanth[2*8], Pramod Kumar Garg[4*9]*

Author's affiliations:

1-Infection and Immunology Division, Translational Health Science and Technology Institute, Faridabad, India

2-Laboratory of Gut Infection and Inflammation Biology, Regional Centre for Biotechnology, Faridabad, India

3-Department of Pediatrics, All India Institute of Medical Sciences, New Delhi

4-Department of Gastroenterology and Human Nutrition, All India Institute of Medical Sciences, New Delhi, India

5-Department of Pathology, All India Institute of Medical Sciences, New Delhi, India

6-Pediatric Biology Center, Translational Health Science and Technology Institute, NCR Biotech Science Cluster, Faridabad, India

7-Department of Surgery, All India Institute of Medical Sciences, New Delhi, India

Footnotes:

 # Equal contribution

8 Senior authors with equal contribution

9 Lead author

*Corresponding authors

Contact info:

Pramod Kumar Garg: *Department of Gastroenterology and Human Nutrition, All India Institute of Medical Sciences, New Delhi -110029, India. Email: pgarg10@gmail.com*

Chittur V Srikanth: *Laboratory of Gut Infection and Inflammation Biology, Regional Centre for Biotechnology, Faridabad – 121001, India. Email: cvsrikanth@rcb.res.in*

Bhabatosh Das: *Functional Genomics Laboratory, Centre for Microbial Research, Infection and Immunology Division, Translational Health Science and Technology Institute, NCR Biotech Science Cluster, 3rd Milestone, Faridabad – Gurgaon Expressway, PO box 04, Faridabad – 121001, Haryana, India.* Phone: +91-129-2876471; Fax: 0129-2876500; *Email: bhabatosh@thsti.res.in*

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Summary:

Gallbladder cancer (GBC) is an aggressive gastrointestinal cancer. Gallbladder stones and chronic microbial infections though implicated in its carcinogenesis are not proven risk factors and the underlying mechanisms are largely unknown. Here, we show an increased abundance of *Salmonella* in the gut microbiome of patients with GBC and culturable *Salmonella* Typhimurium from their GBC tissue. Comparative genomics of *S.* Typhimurium isolated from GBC tissue showed a high invasive index. *S.* Typhimurium isolates harbored horizontally acquired potential virulence functions in their accessory genome. Chronic *S.* Typhimurium infection displayed inflammation, pre-neoplastic changes and tumor promoting mechanisms in a mouse model with gallbladder stones, including activation of the epigenetic modulator Kdm6b. Inhibition of Kdm6B reduced tumor size in NOZ cells engrafted in SCID-mice. Transcriptomic analysis of human GBC tissue supported the proposed mechanisms. Thus, we show a causal association between GBC and chronic *Salmonella* infection and a host epigenetic mechanism in gallbladder carcinogenesis.

Keywords: Gallbladder cancer, Gallstones, *Salmonella*, Microbiome, Infection, Inflammation,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Introduction:

Gallbladder cancer (GBC) is one of the most lethal cancers with a 5-year stage-independent survival rate of 20% and overall mean survival of <6 months (Shaffer & Hundal, 2014). The dismal prognosis is due to aggressive biological behavior of the tumor and the delay in diagnosis (Batra et al., 2005). Even in the United States, <20% of GBC are diagnosed in the early stages. Curative surgery is possible in only 10-20% of patients and the response to chemoradiotherapy is also poor (Sharma et al., 2010). The incidence of GBC is high in many parts of the world such as Chile, Mexico, Bolivia, India, non-Hispanic blacks in the US, around the Mediterranean and Japan (Sharma et al., 2017). For example, the incidence of GBC was reported to be 12.3/100,000 for males and 27.3/100,000 for females in Chile (Shaffer & Hundal, 2014) necessitating the need for population based preventive strategies.

Although various risk factors for GBC have been reported, a definite causal association with a particular risk factor has not yet been established. The most likely causal association has been with gallbladder stones since 60-90% of patients with GBC have associated gallbladder stones (Bolyen et al., 2019; Bruno et al., 2009). Cohort studies in patients with gallstones have also suggested an increased risk of GBC (Attili et al., 1995). Patients with gallstones have been shown to develop pre-neoplastic lesions in the gallbladder mucosa. In a study of 350 patients with gallbladder stones, we have shown that there was sequential occurrence of one or more preneoplastic lesions in the gallbladder in 64% of patients – hyperplasia in 32%, metaplasia in 47.8%, dysplasia in 15.7%, and carcinoma in situ in 0.6% (Jain et al., 2014). In addition, we found accumulation of loss of heterozygosity (LOH) in *p53, p16, DCC*, and *APC* genes in those pre-neoplastic lesions. However, despite evidence of a strong association of gallstones with GBC, contrast between countries with a high and low prevalence of gallstones, and incidence of GBC questions the causal association of gallstones as a single risk factor for GBC. For example, 10-15% of the US adult population have gallbladder stones but the overall incidence of GBC is much lower i.e. 1.3/100,000 (Everhart & Ruhl, 2009; Stinton & Shaffer, 2012). On the other hand, the prevalence of gallstones is lower at 4.1% in India but the age standardized incidence of GBC is much higher at 11.8/100,000 population (National Cancer registry programme 2012-2014) (Horesh et al., 2021). These observations suggest additional risk factor(s) in the etiopathogenesis of GBC. Chronic gallbladder infection has long been suspected to be an important risk factor for GBC. In a case-control study, we had shown that chronic typhoid carrier state was associated with GBC (Odds Ratio = 14; 95% CI 2-92). A meta-analysis of >1000 cases of GBC has shown that the relative risk was 4.6 (95% CI: 3.1-6.8) for anti-*Salmonella* Vi antibody and 5.0 (95% CI: 2.7-9.3) for bile or stool culture positivity for *Salmonella* infection

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

(Koshiol et al., 2016). *Salmonella* is transmitted by the feco-oral route which is facilitated by poor hygiene, and food and water contamination – conditions prevalent in areas with a high incidence of GBC (Jain et al., 2013). Gut microbiota derived toxins and metabolites have been implicated in carcinogenesis of some cancers (Rowland, 2000). However, the role of a particular chronic bacterial infection in the carcinogenesis of GBC and mechanism(s) thereof are not well known. In the present communication, we report results of: (1) human studies to examine (a) alteration in gut microbiome in patients with GBC (b) demonstration of viable *Salmonella* in the gallbladder tissue of patients with GBC, (c) virulence and invasive factors of the isolated *Salmonella* in patients with GBC and (d) genome-wide expression in GBC tissue to understand the mechanistic pathway(s) in its carcinogenesis, and (2) experimental studies to show (a) the role of enteric *Salmonella* infection and gallstones in the carcinogenesis of GBC in a murine model, and (b) the epigenetic alterations in the pathogenesis of GBC.

Results:

Composition of gut microbiota in patients with gallbladder cancer is altered and is enriched with *Salmonella* and other proteobacteria: In a case-control study, we investigated the fecal microbiome in 20 GBC patients and 20 patients with gallbladder stones (GS) and compared it with previously reported 50 healthy adult Indians living in the same geographical area (Das et al., 2018). Demographic details of patients are provided in **Suppl. Table S1**. The community microbial DNA extracted from fecal samples generated a total of 17,52,338 raw reads and 17,17,739 processed reads (post quality filtering and de-multiplexing), with an average 44,459 reads per subject (average varying between 72889 and 22719 reads across the samples). The average read length obtained was 1306 bps, which covers all the variable regions of the16S rRNA gene (read averages ranged from 972 to 1366 bps) **(Table 1, Suppl. Table S2)**. Our analysis showed that the composition and relative abundance of *Actinobacteriota, Bacteroidota, Cyanobacteria, Spirochaetota,* and *Verucomicrobiota* were similar in the fecal samples of GBC and GS patients and healthy subjects (**Figure 1; Suppl. Table S2).** However, the relative abundance of three bacterial phyla viz. *Proteobacteria, Acidobacteria*, and *Tenericutes* was significantly higher in the GBC and GS patients compared to the healthy subjects (**Figure 1)**. Furthermore, we observed distinct interquartile range (IQR) of relative abundance distribution in *Proteobacteria* (IQR_{GBC} = 0.0232, IQR_{GS} = 0.0122), *Acidobacteriota* (IQR_{GBC} = 0.0126, IQR_{GS} = 0.00904), and *Tenericutes* (IQR_{GBC} = 0.00899, IQR_{GS} = 0.00621) in the fecal samples of GBC and GS patients compared to the healthy subjects. Among all three phyla, the IQR of relative abundance of the phylum *Proteobacteria* was considerably higher in the GBC group (**Figure 1).**

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

The composition of the fecal microbiota was further examined at the levels of class, order, genus and species. The sequencing depth employed here revealed a total of 1001 bacterial genera and 505 different species in all the GBC and GS patients. On analyzing the alpha diversity of the fecal microbiota of GBC, GS patients and healthy subjects, we found that there was a significant difference in the Chao index (Kruskal Wallis H test, P value =5.5e-0.8), whereas no difference was observed in Shannon and Simpson diversity indices **(Suppl. Figure S1; Suppl. Table S3)**.

To identify disease specific microbial taxa, we considered the differences in relative abundance of bacterial taxa. At the genus level, we observed that 29 genera were differently abundant across the three groups (one-way ANNOVA at P-value<0.05). The relative abundance of *Salmonella,* was higher in GBC compared to GS patients. The highest abundance of *Salmonella* observed in a GBS patient was 0. 0226. Interestingly, *Salmonella* was absent in the fecal sample of any healthy subjects. The genus *Streptococcus* was significantly high (Kruskal Wallis H test, P-value<0.05) in the GBC patients compared to the GS and healthy subjects **Suppl. Figure S2**. However, a few of the healthy samples were also found to have *Streptococcus* in their fecal samples. The *Prevotella pallens* had lower abundance in GBC patients as compared to GS, whereas it was not detected in the healthy subjects. We also observed that 7 bacterial genera belonging to *Proteobacteria* had significantly different abundance patterns across the three groups (Kruskal Wallis H test, P-value<0.05) **Suppl. Figure S2**. *Salmonella* (p-valve: $5.22x10^{-18}$), *Burkholderia (*p-value: $5.137x10^{-18}$*), Serratia* (p-value: $5.250x10^{-18}$), *Acinetobacter (*p-value: $3.306x10^{-14}$*), Klebsiella* (p-value: $1.34x10^{-8}$)*, Pseudomonas* (p-value: $2.01x10^{-7}$) and *Haemophilus* (p-value: $2.11x10^{-6}$)*,* were observed to have the higher abundance in GBC and GS patients as compared to the healthy subjects (**Suppl. Figure S2; Suppl. Table S4).** The relative abundance of *Salmonella, Burkholderia* and *Serratia* was higher in GBC compared to GS patients. Furthermore, the genera *Salmonella*, *Serratia* and *Burkholderia* were absent in healthy subjects.

Culture independent analysis of microbial composition in the GBC tissue:

Since we observed that the relative abundance of *Salmonella* in the gut microbiota of GBC patients was very high compared to the other two groups, we studied if *Salmonella*, being invasive in nature with an ability to survive in bile, was present in the GB cancer tissue. We tested the microbiota composition of surgically resected gallbladder cancer tissue from the GBC patients (n=17). Of these 17 samples, we could amplify the 16S rRNA region of the bacteria in 9 samples. After quality filtration ~72% to 76% paired reads of 16S rRNA gene derived from the gallbladder tissue were subsisted for taxonomic assignment **(Suppl. Table S2)**. About 26% to 34% quality filtered paired reads with average of ~435-bp read lengths were used as input to the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

QIIME2 program to generate the OTU table. The OTUs rarefaction curve showed very less microbial diversity and around 30 OTUs were identified (**Suppl. Figure S3)** which suggested that the gallbladder might not be a conducive residence and could provide a niche environment for only a few microbial species. The identified 30 OTUs belonged to 7 phyla in which *Proteobacteria* was the most dominant **(Suppl. Figure S4)**. The numbers of reads of the remaining phyla were too less that they could be considered as a rare phylum of the ecological niche. At the genus level, *Salmonella* was the most dominant bacterium across all 9 samples **(Suppl. Figure S5)**. Apart from *Salmonella*, four other bacterial genera namely *Pseudomonas*, *Burkholderia*, *Klebsiella*, and *Ralstonia* were also observed in most of the gallbladder tissue samples with similar abundances.

Insights into genomic potency and functionality of *Salmonella* residing in the gallbladder tissue of GBC patients:

Since *Salmonella* was the most abundant bacterial genus and distributed in all 9 GBC tissue samples, we isolated the bacterium for exploring its genomic repertoires, and genetic potential with regard to virulence and disease pathogenesis. The characterization of the taxonomy of the isolate by complete 16S rRNA gene sequencing revealed it to be *Salmonella enterica* var Typhimurium. Whole genome shotgun sequencing of 28 *Salmonella* Typhimurium isolates was performed. A total of 11974408 paired-end (PE) reads (~6.71GB data) with 280 bp mean read length were generated. Of these reads, 11862942 (~6.64 GB data) had Phred Quality Score (Q) >20 and were used by SPAdes to generate 11x to 43x genome coverage (**Suppl. Table S5.1**). RAST annotation details are mentioned in **Suppl. Table S5.2.**

Genomic potency and diversity of the *S.* Typhimurium were determined based on the MLST sequence types (ST) and also by whole genome SNP analysis. Among the study isolates, 23/28 were identified as ST19, while 5 genomes were of unknown STs due to one or two missing genes in the draft assemblies. However, the available allelic profiles of 5/7 genes were found similar to ST19 and thus we predicted these genomes belonged to clonal complex ST19 (CC19) (**Table 3**). Further, whole genome SNP based phylogenetic analysis of *S.* Typhimurium was performed along with the previously published genomes belonging to ST19, in comparison to the other most common STs of ST34 and ST313 collected across the globe (*n=2490*) (**Suppl. Table S6)**. Majority of these were sourced from fecal samples, except the African strains of invasive ST313. The phylogenetic positioning of *S.* Typhimurium was characterized in regards to their STs and spatiotemporal profile. The GBC sourced isolates in the present study clustered within

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

the ST19 genomes (**Figure 2a**). Notably, within the ST19 group, the contemporary GBC isolates formed a distinct sub-cluster along with 12 other genomes from North America and Europe sourced from fecal samples during the period 2006 to 2020 (**Figure 2a**). Five isolates that were not genotyped by MLST also clustered within the study isolates of ST19, confirming these genomes as ST19/CC19.

Invasive index of ST19 *S.* Typhimurium isolated from the GBC tissue

The potential invasiveness of each of the GBC sourced ST19 *S.* Typhimurium isolates was investigated against the genome degradation event captured with the pre-defined 196 genes. The GBC isolates displayed an invasiveness index score of predominantly >0.0823 (except for a few), which was slightly higher than reference *S.* Typhimurium LT2 (0.0823) (**Figure 3a).** Minor differences in the score could be due to the limited genome degradation events with a few genes out of the 196 genes in the GBC isolates as compared to the African invasive ST313. Of note, 1/3rd of isolates from each patient displayed a higher invasive score compared to the other isolates sourced from the same patient, indicating heterogeneity and the likelihood of within-host evolution of *S.* Typhimurium. On comparison, we found that the isolates with a high mutation frequency displayed a high invasive index score (except for P7B, P8B and P9B) (**Figure 3b)**. Thus, the overall observed signatures of gene degradation events led to the increased invasive potential of this pathogen in GBC than the non-invasive ST19 strains.

As the invasive isolates in the present study were sourced from GBC, the dynamics of pseudogenization/gene degradation that favored the invasiveness were examined to identify host adaptation signatures in comparison with the African invasive ST313. This phenomenon of hypothetically disturbed coding sequences (HDCS)/pseudogenes has been characterized by the presence of multiple mutations and was captured in the GBC isolates mapped against the reference genome *S.* Typhimurium LT2. Firstly, all the genomes displayed a notable 10 bp nucleotide deletion in GBC ST19 isolates causing a frameshift mutation in the virulence gene *lpfD.* This encodes for a long polar fimbriae that helps colonize the Peyer's patches but its inactivation due to frameshift prevents the pathogen's interaction with M-cells. Further, in the mutational profiling, missense mutations were the most abundant that ranged from 193 to 261 mutations among all GBC *S.* Typhimurium isolates (**Figure 3b)**. In addition, 11 frameshift mutations were found to be common in all the 28 genomes, while a few genomes harbored additional mutations that ranged from 11 to 18. The mutation profile was compared with the top 10 invasive ST313 signatures to infer the invasiveness potential of the GBC isolates **Suppl. Figure S6a**. First notable signature, a one-bp deletion in the *ratB* gene (STM2514: 5807delT), resulted in the frameshift of the protein at position 1936/2435, thereby forming a pseudogene.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Second was the missense mutation in the *sseI* gene encoding for Gifsy-2 prophage putative type III secreted protein (STM1051: T308C) which led to the amino acid change of Asp103Gly (D103G), making this as a non-functional protein. Third was the inactivation of a virulence gene *avrA* (one nucleotide deletion 799A), an effector protein that inhibits the NFκB signalling pathway. Its non-functionality hinders with the modulating ability of the pro-inflammatory response during an infection. These three key signature changes might have favored the ST19 *S.* Typhimurium in crossing the intestinal barrier to become invasive. Further, 18 genes with missense mutations that were predefined as pseudogenes in the ST313 population were also found in ST19 GBC isolates. Eleven other frameshift mutations were also identified besides ST313 signatures. Moreover, the differences in the mutational profile (missense) within the 28 study isolates were observed with 3 strains each from 8 patients and 2 strains each from 2 patients, respectively **Suppl. Figure S6b**. Such heterogeneity in the mutational patterns within the same patient is remarkable and is being documented for the first time.

_S._ Typhimurium isolated from the GBC harbored distinct accessory genomes containing potential virulence functions:

A comparative pan genome analysis was carried out to analyze the differences in the genome composition of different STs such as ST19, ST34 and ST313. Among the 2518 genomes, we identified 13,750 as total gene families. The number of core genes (99 to 100%) was 3822, soft-core genes (95 to 99%) were 185, shell genes (15 to 95%) were 1123, and cloud genes were 8620 (0 to 15%). The phylogenetic tree with the clades of each represented STs showed similar gene content profiles. However, regions/genes that were likely to be shared by the specific STs were highlighted as shown in the Phandango illustration **(Figure 4)**. These main differences were in the accessory genome content with the presence of prophage elements that were also ST specific. **Figure 4a-d** shows the complete pan genome profile of the *S.* Typhimurium with respect to different STs and their associated metadata.

Putative virulence factors screened by the VFDB analyser tool identified 11 virulence factor classes corresponding to 160 virulence factor genes in the study isolates. The candidate virulence factors are associated with secretion systems which include TTSS (SPI01encode), TTSS (SPI02encode), TTSS effectors translocated via both systems, TTSS01 translocated effectors and TTSS02 translocated effectors. Additionally, VF class like fimbrial adherence determinants (*csg*, *bcf*, *fim*,*lpf*, *pef*,*saf*, *stb*, *stc*, *std*, *stf*, *sti*, *stj*) macrophage inducible genes (*mig-14*, *mig-5*), magnesium uptake (Mg^{2+} transport), non fimbrial adherence determinants (*misL*, *ratB*, *shdA*, *sinH*), serum resistance (*rck*), and stress adaptation (*sodCl*) genes were also identified in the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

GBC isolates . The different VF classes along with their corresponding VF genes found in the present study isolates are illustrated in **Suppl. Figure S7**.

Three important proinflammatory SPI-1-T3SS effector proteins *sopE2*, *sopB/sigD* and *sopA* were found in all 28 isolates. In contrast, one of the T3SS effector protein *sopD* was present in 22/28 (78.5%) genomes. Virulence factor genes with anti-inflammatory properties, *gogB*, *gogA*, *SseK1*, *SseK2*, *SopA*, *SopB/sigD, sspH2* were identified in all the 28 genomes whereas *sptP* and *spvC* were identified in 11/28 (39%) and 12/28(42%) genomes respectively. Other effector proteins like, *spvD*, *spvC* and *avrA* which directly target inflammatory signalling in a negative way were also found to be present in the GBC *Salmonella* isolates. Of these, *spvD*, which is involved in inhibiting NF-κB (Rolhion et el., 2016) was identified only in 2 (0.7%) genomes (P5C and P6B). *sodCI* which corresponds to stress adaptation was found to be present in 96% (27/28) of genomes. The presence and absence of these pro-inflammatory and anti-inflammatory virulence genes in the GBC isolates are depicted in **Suppl. Figure S8**. From our analysis, we found that the same virulence profile was seen in the isolates from four patients whereas there was heterogenicity in the distribution of virulence genes in isolates from five patients.

Horizontally acquired plasmids and prophages with virulence potency in the genome of *S.* Typhimurium isolated from the GBC tissue:

Plasmids harboring virulence genes are of special importance since they are involved in the dissemination of antibiotic resistance. PLSDB identified four circular plasmids in the GBC *Salmonella* isolates **(Table 4)**. First was a 69.65 Kb *S.* Typhimurium strain AUSMDU00027951 plasmid P02 (NZ_OU015330.1) with a GC content of 52.67% in 28 sequenced genomes. Second was a 147Kb *Salmonella enterica* subsp. Senftenberg strain NCTC10384 plasmid 3 (NZ_LN868945.1) has a similar GC% as the previous plasmid in 15 of the 28 (53%) genomes. Both these plasmids harbor *entD*, which encodes phosphopantetheinyl transferase component of enterobactin synthase multienzyme complex (Enterobactin) as the virulence factor. Third plasmid was a *Salmonella enterica* subsp Enteritidis strain 81-1705 plasmid pSE81-1705-3 (NZ_CP018654.1) of 33Kb in 96% (27/28) of the genomes of *S.* Typhimurium, which harbored *galU*, a glucosephosphate uridyl transferase gene involved in immune modulation. The fourth plasmid was present in 39% of the GBC isolates. This plasmid carries 5 virulence genes (*pefC*, *spvB*, *spvC*, *rck*, *pefB*). Besides, 2 plasmid replicons having a capacity to activate and control replication (*InCFII* and *InCFI*) were also identified within this plasmid. On basic local alignment search tool BLASTn analysis, we found that *Salmonella enterica* subsp. CFSAN002003 plasmid showed identity and coverage of 99.99% with the virulence plasmid of *S.* Typhimurium LT2

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

(pSLT) which is of 94Kb. pSLT has an 8Kb region which encodes the virulence gene *spv* which has been reported to increase the bacterial growth rate at the time of the systemic phase of the disease progression (Hiley et al., 2019). Heterogenicity was observed in 60% of the patients whereas in only 40% of the patients there was similarity within the isolates. BLASTn results further confirmed the presence of these four plasmids within the GBC isolates having query coverage of >90 and percent identity of >98%.

Three lysogenic prophages were initially identified by the Phage search tool enhanced release PHASTER in the annotated genome of *S.* Typhimurium. Gifsy-1, Gifsy-2 and salmon118970_sal3 were commonly present in all the 28 *S.* Typhimurium isolates. Salmon118970_sal3 possesses one virulence gene (*SseK2*). Gifsy-1 (encoding *gogA* and *gogB*) and Gifsy-2 (encoding *sodCl* and *ssel*) along with four phage remnants Def1-4 having virulence genes were also identified in the genome of ST313 representative strain D23580 which has been reported to cause invasive disease in sub-Saharan Africa (Kingsley et al., 2009). We then did a manual BLASTn analysis to confirm the invasiveness in our studied GBC isolates against *S.* Typhimurium D23580 genome as a reference. Results indicated that Def-2 encoding *sopE2* as the virulence gene was found in all the 28 genomes. Def-1 was also found to be present in the three of the genomes with a query coverage of less than <35%. Three GBC genomes of *S.* Typhimurium were found to possess Def-3 remnants having an anti-inflammatory virulence gene *sspH2* with less query coverage in comparison to the other isolates. A 20Kb region (*gtrBb*) from the Def-4 phage encoding virulence trait was identified in 25 GBC isolates. ST64B prophage with the virulence gene *SseK3* was identified in all 28 isolates. The overall profile for the phage analysis within the isolates was found to be similar. The presence and absence of prophage and prophage remnants in the GBC isolates in comparison to D23580 are mentioned in Table 5.

Enteric *Salmonella* chronically colonizes Gallbladder in a Gallstone mice model

To examine the role of chronic *Salmonella* infection in the pathogenesis of GBC, we developed a modified version of a mice model of gallstone disease along with chronic *Salmonella* infection following a previously published protocol (Crawford et al. 2010). Our methodology involved four weeks old wild type FVB mice (4-6 animals per cohort) fed with mouse chow supplemented with 1% cholesterol and 0.5% cholic acid for 9 weeks. As shown in **Figure 5**, lithogenic diet feeding led to development of gallstones in these mice (hereafter referred as Lithogenic Diet or LD group). A subset of gallstone mice was infected with *Salmonella* typhimurium (strain SL3261) by oral gavage (referred to as Lithogenic diet + ST

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

group or LDST). The SL3261 is an attenuated *Salmonella* strain bearing an aroA deletion well suited for chronic infection in mice (Eisele NA et. al. 2013). Separate groups of mice fed with a regular diet and mock-infected (Control) or those fed with a regular diet but infected with SL3261 (Control ST) were also included in the study.

Gallstone developed in 100% of the lithogenic diet fed animals. The stones were multiple with a range of different colors including dark green, red (in LD-ST group) and pale yellow (in the LD group). The size of the stones ranged approximately between 0.1-0.5mm in diameter (**Figure 5b**). Mice infected with *Salmonella* also developed splenomegaly (**Figure 5c**). In these mice (LD group and LD-ST group), the liver was enlarged with a change in color to pale red possibly due to diet induced accumulation of fat (LD group and LD-ST group), while the control mice did not show any of these phenotypes (**Figure 5**c). Notably, the LD-ST group mice showed the presence of *Salmonella* in internal organs, 15 days post infection, as revealed by colony forming units (cfu) of *Salmonella* in the liver (**Suppl. Figure S9**), and the portion of the common bile duct connected to the gallbladder. The presence of *Salmonella* was further confirmed by 16S qRT-PCR assay and electrophoresis of the amplicons on 1.5% agarose gel (**Figure 5d**). *Salmonella* fecal shedding was also seen by the presence of its colony forming units (cfu) in the MacConkey agar plates and further confirmed by qRT-PCR assay and agarose gel electrophoresis. (**Figure 5**e). Together these data revealed the successful development of a gallstone-mice model with chronic *Salmonella* infection.

Chronic Salmonella infection and gallstone disease cause chronic inflammation and induce pre-malignant changes in the gallbladder mucosa

The histology of the gallbladder tissue sections was analysed. As anticipated, the control group showed healthy normal tissue histology (**Figure 6a**). However, signs of inflammation marked by infiltration of immunocytes neutrophils were seen in the LD-group, ST-group and in LD-ST mice. Importantly, pre-malignant changes in the form of hyperplasia and metaplasia developed only in the LD-ST group at 4 months post infection (**Figure 6a**). These changes were also evident in the histopathology scoring (**Figure 6a** right panel). These pre-malignant changes were not evident at 15 days post infection in these groups. At later stages of infection (i.e. 8.5 month post infection), more severe changes such as thickening of the gallbladder wall, fibrosis, hyperplasia and metaplasia were evident in the LD-ST group (**Figure 6c**). The pathological scores indicated that wall thickening and hyperplasia were maximum in LD-ST group followed by Control-ST mice. Taken together, these data led us to conclude that gallstone along with chronic *Salmonella* infection caused chronic inflammation

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

and induced pre-malignant changes in the gallbladder tissue. Mucins are a critical determinant of the nature of a tissue and are often used as an indicator of different pathological conditions. Neutral mucins are present in healthy gallbladder tissue epithelium, whereas acidic mucins are present in metaplasia associated gallbladder tissues. Therefore, we carried out AB-PAS staining to examine the type of mucins in the different groups of mice. The gallbladder tissue sections of 4 months infected LD-ST group and LD-group mice showed multiple zones of deep blue staining which indicated the presence of acidic mucins (**Figure 6c**). In contrast, the control and Control-ST group showed normal neutral mucins. These results are indicative of significant mucosal changes in the LD-ST mice (**Figure 6c**).

Chronic Salmonella infection induces epigenetic changes involving Kdm6B in the murine gallbladder

We recently reported the involvement of Kdm6B, a histone demethylase, to be necessary for chronic *Salmonella* infection (Rana et al., 2021). In this report, up regulation of Kdm6B expression was shown to reprogram infected macrophages into non-bactericidal M2 like spectrum (Rana et al., 2021). The presence of pre-malignant changes in the LD-ST mice tissue sections described above prompted us to investigate possible involvement of epigenetic alterations. We examined the status of Kdm6B in the *Salmonella* infected gallstone-mice. We found that the expression of epigenetic modulator Kdm6B was upregulated at the mRNA level in LD-ST mice compared to the other groups (**Suppl. Figure S10a**). At the protein level as well, Kdm6B expression was higher in *Salmonella* infected mice (**Figure 7a-b**) at 5 month post infection. To see the effect of infection in longer durations, we allowed infection to proceed for 1.8 years. The increase in Kdm6B was dramatic in LD-ST mice (2 out of 4 mice) at 1.8 years (**Figure 7b**). Taken together, these results suggest that chronic *Salmonella* infection alone increased Kdm6B expression but along with presence of gallstones led to further increase in the Kdm6B expression which could contribute to the formation of pre-malignant lesions.

Therefore, we examined if *Salmonella* induced Kdm6B upregulation may be contributing to activation of genes favoring tumorigenesis. A decreased expression of tumor suppressor genes p16 and p53 was observed in Control-ST and LD-ST group compared to the control group and LD group. p16 showed significant ~5-fold downregulation in all the 3 groups (LD group, ST group and LD-ST group mice) compared to uninfected control of 8.5 months batch (**Figure 7c-d**). p53 was also drastically downregulated in gallbladders of LD, LD-ST and Control-ST group but not in untreated control group of mice of 8.5 months post infection.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Notably, among these groups, ST mice showed 13-fold downregulation compared to Gallstone-ST, which showed 10-fold decrease in expression (**Suppl. Figure S10**). These data indicate pro-tumorigenic molecular alterations occurring in gallbladder tissue of Gallstone-ST and ST mice.

Kdm6B in Human GBC:

The status of Kdm6B in human gallbladder cancer tissue was examined by immunoblotting from lysates of patient biopsy samples (**Figure 7e**). A significant increase in Kdm6B expression in GBC samples compared to those from gallstone or non-malignant patient controls was observed. In the case of gallstone samples, only a slight upregulation of the expression of Kdm6B was seen.

Genome-wide expression profiling of human gallbladder cancer to understand the pathways involved in GBC:

Genome-wide expression profiling was done in the gallbladder tissue samples collected from 16 GBC patients and 16 gallstones controls (matched for age and sex) that underwent cholecystectomy. All 16 were adenocarcinoma on histology. The age and sex distribution of the patients and controls are provided in **Suppl. Table S7**.

Differentially expressed genes on mRNA profiling:

To identify DEGs associated with GBC, expression profiles of 32 human samples including 16 adenocarcinoma patients and 16 samples from gallstone disease controls were analyzed using LIMMA. The criteria of FC $\geq$1.5 for up-regulated and FC<1.5 for down-regulated with $\leq$ 0.05 were used to identify the significant DEGs. The total number of differentially expressed transcripts was 480 that had significant p-value and adjusted p-value (**Figure 8**). Overall, 340 genes could be annotated from 480 transcripts (205 upregulated and 135 downregulated). The human transcriptomics data suggested that several genes including ADAMTSL5, CX3CR1, HMGA2, MMP7, EPHB2, TYMP, HLA-A, MK167, and HDGF were differentially regulated in GBC tissue as compared to those with gallstones. The list of differentially expressed genes is given in **Suppl. Table S7.** Functional analysis through KEGG (Kyoto Encyclopaedia of Genes and Genomes) showed that differentially expressed genes were related to metabolic pathways, neutrophil extracellular trap formation, alcoholism and transcriptional mis-regulation in cancer (**Figure 8).** A few of the dysregulated genes mentioned above identified through human transcriptome data were further tested experimentally in the context of *Salmonella* infection.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Experimental evidence of the role of Kdm6B:

The expression of genes, which were identified by GBC tissue transcriptomics, was examined in NOZ cells, a cell line derived from human gallbladder carcinoma tissue. NOZ were cultured and either sham infected or infected with *Salmonella* for 1 hr. Among these genes, those with regulatory function, such as- ADAMTSL5 (maintains oncogenic signaling, master regulator of carcinogenesis), CX3CR1(C-X3-C motif chemokine receptor 1, known for cell migration and tumor invasion) and SPSB4 (regulator of protein ubiquitylation, plays a role in cell invasion) were significantly dysregulated (**Figure 7f**). We next examined if Kdm6B function was required for the upregulation of these genes by carrying out chromatin immunoprecipitation. A 3-fold higher binding of Kdm6B in the promoter of ADAMTSL5 was seen in NOZ cells infected with *Salmonella* compared to control cells (**Figure 7g**). In line with this we also observed a significant decrease in H3K27me3 mark, i.e. histone-3 lysine 27 tri-methylation (**Figure 7h**). Based on background literature and these findings, we hypothesized that *Salmonella* in the gallbladder causes epigenetic alterations to drive tumorigenesis.

Furthermore, infection of NOZ cells resulted in an increase in expression of Kdm6B at 4, 7 and 18 hrs post infection **(Suppl. Figure S11a).** Tumour suppressor gene p53 was also found to be downregulated during infection in a time dependent manner **(Suppl. Figure S11b)**. *Salmonella* specific inflammatory markers IL8 and IL1β were also found to be upregulated. Moreover, expression of IL-8 was increased early during infection i.e. 4 hrs, while IL-1β at later stage of 18 hrs post infection. The protein levels of Kdm6B showed an increase at 4 and 8 hrs post infection in these cells **(Suppl. Figure S11c)**. To further understand the underlying mechanism, we cultured primary gallbladder epithelial cells from mice gallbladder followed by infecting them with mCherry-labeled *Salmonella* (Verma et al., 2015) **(Suppl. Figure S11d)**. At 7-hours post infection, we observed *Salmonella* in the perinuclear region of these cells. The expression of IL-8 was also seen to go up in these cells which is one of the hallmarks of *Salmonella* infection. Interestingly, a concordant increase in the expression of Kdm6B **(Suppl. Figure S11e**) was also observed.

Activation of Kdm6B is known to regulate a large set of genes and thereby modulate fundamental programs of a cell including proliferation and differentiation. To investigate such a role of Kdm6B role in bestowing tumor related properties in gallbladder cells, we tested cell proliferation and cell migratory properties. We experimentally perturbed Kdm6B expression in NOZ cells to examine this. Plasmids encoding Kdm6B or its catalytic mutant

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

(KDM6B^{H1390A}) were transfected into NOZ cells followed by cell proliferation assay. Successful transfection of constructs in NOZ cells was confirmed by RT-PCR (**Figure 6B**). We observed that cells overexpressing wild-type Kdm6B showed a >2-fold increase in cell proliferation, while no change was observed in case of Catalytic mutant (CM) of Kdm6B (**Suppl. Figure S12**). Next, we carried out siRNA mediated knockdown of Kdm6B which also resulted in decreased cell proliferation compared to scrambled siRNA transfected control (**Suppl. Figure S12**). Next, treatment with demethylase inhibitor, GSKJ4, at 15uM and 30uM resulted in decreased cell proliferation in a concentration dependent manner (**Suppl. Figure S12**). In line with this, wound healing assay revealed a higher rate of wound closure in case of cells treated with wild-type Kdm6B compared to its CM at all the time points i.e. 12, 24 and 32 hrs (**Suppl. Figure S12**) a phenotype that was also seen upon GSKJ4 treatment in comparison to the respective controls at 12, 24 and 32 hrs post treatment (**Suppl. Figure S12**).

Next, a sub-cutaneous tumour model with NOZ cells was used to examine the role of Kdm6B in gallbladder tumorigenesis. The NOZ cells were injected subcutaneously into NOD/SCID mice followed by intraperitoneal injection of GSKJ4 (**Suppl. Figure S13**). Post-xenografting, the growth of the xenograft in the control mice showed a steady increase. However, in contrast, GSKJ4 treatment resulted in regression of tumor growth as evidenced by animal body weight, tumour volume and weight (**Suppl. Figure S13**). Taken together these data strongly indicate initiation of pro-tumorigenic mechanism in a Kdm6B dependent manner during *Salmonella* infection.

Discussion:

Carcinogenesis is a multifactorial process involving primarily genetic susceptibility and environmental factors. However, genetic and environmental factors may not account for the entire cancer risk and therefore chronic microbial infection might be the 3rd important co-factor, referred to as a tripartite multidimensional interaction by the International Cancer Microbiome Consortium (Scott et al, 2019). Many chronic microbial infections such as the human papilloma virus, the hepatitis B virus, and *Helicobacter pylori* infection have been shown to be definite carcinogens (Masrour-roudsari et al. 2017). In the case of gallbladder cancer, chronic *Salmonella* infection has also been suggested as a co-factor in its carcinogenesis (Iyer et al., 2016; Scanu et al., 2015; Shukla et al., 2000; Upadhayay et al., 2022). Chronic Salmonella carrier state has been associated with GBC in many historical studies, including more recent studies based on Salmonella Vi antibodies (Welton et al, 1979;

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Mellemgaard et al, 1988; Dutta et al, 2000). A meta-analysis of >1000 patients with GBC showed a relative risk of 4.6 (95% CI: 3.1–6.8) for anti-Vi antibodies positive individuals (Koshiol et al, 2016). However, Vi antibody seropositivity may not reflect Salmonella infection in the gallbladder. Detection of Salmonella in the GBC tissue is difficult by the culture technique and therefore the association remains weak. Furthermore, whether the association of chronic Salmonella infection with GBC is causal or not has not been proven. Does Salmonella reside in the gallbladder in a state called 'dynamic symbiosis' with cancer occurring mainly due to random DNA replication errors (Tomasetti et al. 2017) or the chronic *Salmonella* infection contributes significantly to carcinogenesis and requires robust association and mechanistic studies.

In the present study, we looked at the association of chronic *Salmonella* infection with GBC in human studies and developed a murine model to get a mechanistic insight into the pathogenesis of how chronic infection might contribute to GBC.

In the case-control study, we found that *Salmonella enterica* was present in the fecal microbiome of patients with GBC whereas it was absent in healthy controls. This has not been reported previously. We then studied whether *Salmonella* was present in the GBC tissue. Using a 16S rRNA based targeted metagenomic approach, we found the presence of *S.* Typhimurium in the gallbladder cancer tissue. A few studies have earlier shown the presence of *Salmonella* typhi in the GBC tissue (Pratap et al 2013; Scanu et al 2015; Nath et al 2008). Notably, detection of the non-typhoidal *Salmonellae* was also reported (Iyer et al., 2016). The presence of microbial infection, both bacteria and fungus, within the cancer tissue has been linked with carcinogenesis e.g. *Malassezia* spp. in pancreatic cancer (Thomas et al. 2018; Aykut et al, 2019). *Salmonella* may reside for long in the gallbladder through a biofilm formation and thus contribute to GBC (Di Domenico et al, 2017).

S. Typhimurium being a non-typhoidal serovar, has been predominantly causing gastro-intestinal infections worldwide. Over the past decade, invasive non-typhoidal *Salmonella* (iNTS) associated blood stream infections have been reported widely in Africa, with two-thirds due to *S.* Typhimurium (Marchello et al., 2019). These were the multidrug resistant phenotypes that belonged to the sequence type of ST313, that differed from ST19 by 700 SNPs (Kingsley et al., 2009; Okoro et al., 2012, Feasey et al., 2015). In India, ST19 was the most common iNTS serovar, while a single case of iNTS ST313 was also reported (Jacob et al., 2019, Jacob et al., 2020). To the best of our knowledge, here we report for the first time the invasive ST19 *S.* Typhimurium that was associated with gallbladder cancer in the Indian patients.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

The genomes of *S.* Typhimurium isolated from the gallbladder tissue have been analyzed in detail for the host adaptation genomic signatures, as ST19 were predominantly associated with gastrointestinal diseases. We identified certain key signatures that could have led this pathogen to become invasive in crossing the intestinal mucosal surface. Among the top 10 signatures that led ST313 to become more invasive, eight genes were conserved in all the 28 GBC isolates (Pulford et al., 2021) whereas, two pseudogenes in ST313 *ratB* and *sseI* were also found to be inactivated in GBC isolates. Inactivation of *ratB* reduces the enteric potential and *sseI* increases the rapid dissemination ability of the bacteria to the lymph nodes from the gut (*Okoro et al., 2015; Carden et al., 2017*).

S. Typhimurium harbors effector proteins of SPI-1 and SPI-2 T3SS, which are known to induce proinflammatory signalling and trigger inflammation (Bruno et al., 2009*).* The virulence of *S*. Typhimurium was shown to be conferred by two inducible prophages (Gifsy1-2) (Figueroa-Bossi N et al., 1999). In the present study, the GBC tissue isolates contained mainly three functional prophage sequences (Gifsy1-2, STM64) and four phage remnants (Def1-4) having virulence genes encoded in them. *gogB* from Gifsy-1 phage has a protease activity by targeting the NF-κB transcription factors RELA and RELB, thereby reducing the inflammatory response to *S.* Typhimurium. It is believed that this anti-inflammatory property will reduce tissue damage during long-term infection, whereas inflammation that lasts for a short time may enhance colonization in the intestine (Pilar et al., 2012).

At least two virulence genes have been reported to be carried by Gifsy-2 phage. The first being *sodCl* and the second is an unidentified factor(s). *sodCl* is responsible for encoding one of the two periplasmic Cu/Zn superoxide dismutases of bacteria *Salmonella (*Farrant et al., 1997, De Groote et al., 1997*).* These phage and phage remnants are known to play an important role in infection (Stanley et al., 2000, Figueroa-Bossi N et al., 2001). Almost 88% of *S*. Typhimurium are reported to carry a 90Kb virulence plasmid (Helmuth et al., 1985). In our study, four virulence plasmids were identified in the GBC tissue isolates which carry known virulence genes. PlasmidP02 and plasmid3 carried *entD* (enterobactin) as the virulence gene whereas CFSAN002003 plasmid, also a virulence plasmid, is encoded by *pefC, spvB, spvC , pefD, rck, pefA, pefB* virulence factor genes. The molecular functions of *spvB, spvC* are not clearly known although *spvB* is known to be involved in secretion (Gulig et al., 1993). Another region on the plasmid included a plasmid encoded fimbriae loci known as *pef*, which is involved in adhesion of bacteria to epithelial cells of the intestine (Baumler et al., 1999) and lastly a *rck* gene responsible for resistance to complement killing (Heffernan et al., 1992).

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

With regard to the mechanistic insight, we found that gallstones induced chronic inflammation in the murine model, but when these mice were infected with *Salmonella*, they developed significantly greater chronic inflammation in the gallbladder. In the chronic model when the infection was prolonged, the GB epithelium developed pre-malignant changes. These findings are similar to our previous observations in human gallbladder stones disease, where gallstones-induced chronic inflammation also led to development of pre-neoplastic lesions (Jain et al., 2014).

The molecular mechanisms through which microbial infection may be involved in carcinogenesis are complex and not well understood except in a few instances e.g. HPV and HBV infection. Chronic inflammation, immune interactions, and genotoxicity may be the primary drivers of carcinogenesis.

Inflammation is at the core of carcinogenesis, underlying many microbial associations with cancer e.g. *H. pylori* and gastric cancer (Francescone R et al., 2014). Microbial virulence factors induce chronic inflammation in the host tissue, stimulating cellular proliferation. Pathogen-associated molecular patterns (PAMPs) induce proinflammatory effects by acting on host pattern recognition receptors, such as toll-like (TLR) and nucleotide-binding oligomerization domain-like (NOD) receptors. Downstream activation of cell survival pathways may contribute to carcinogenesis even remotely (Dapito DH et al, 2012).

In the murine gallbladder, significant inflammation was induced by the chronic *Salmonella* infection in the present study. Three important proinflammatory SPI-1-T3SS effector proteins sopE2, sopB/sigD and sopA were found in *S.* Typhimurium by others (Hapfelmeier S et al., 2004; Zhang Y et al., 2006; Thien K et al., 2007). Immune modulation also contributes to the persistence of infection and chronic inflammation. *S.* Typhimurium, which harbours *galU*, a glucosephosphate uridyl transferase gene may cause immune modulation.

Many bacteria produce toxins, which might cause DNA, damage e.g. Cytolethal distending toxin (CDT) and colibactin produced by *Escherichia coli* and Campylobacter jejuni induce DNA breaks (N esić D et al.,, 2004). *S.* Typhi has been shown to produce such CDT. At the molecular level mostly three genes are responsible for the formation of cytolethal toxins i.e., CdtA, CdtB, and CdtC. CdtB gene is responsible for carcinogenic activity and creates the double-stranded DNA break in host cell DNA (Upadhayay et al., 2022). But *S.* Typhimurium lacks the gene for these toxins. Instead, it produces another toxin AvrA which has been shown to be involved in functional modulation of host *p53*. The significance of this modification in the context of gallbladder cancer is not fully known and needs future investigation. Rck, another protein encoded by *Salmonella* Typhimurium was shown to cause

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

DNA double stranded breaks thereby imparting cyclomodulin with a genotoxic activity (Mambu J et al., 2020).

In an earlier study, we showed that chronic *Salmonella* infection could induce epigenetic changes in the murine model. The involvement of Kdm6B, a histone demethylase, is necessary for a successful chronic *Salmonella* infection. In the present study, we also demonstrate that histone demethylase KDM6B recruitment on the PPARδ promoter results in the loss of the H3K27me3 mark leading to its transcriptional activation. PPARδ is a fatty acid oxidation regulator that modulates macrophage polarization, favoring the M2 spectrum and resulting in *Salmonella* chronic infection (Rana et al., 2021). Our human transcriptomics data suggest that several genes such as *ADAMTSL5*, *CX3CR1, HMGA2, MMP7, EPHB2, TYMP, HLA-A, MK167, HDGF* are differentially regulated in the tumor tissue samples from gallbladder cancer patients as compared to the normal epithelium in patients with gallstones. Functional analysis through KEGG showed that differentially expressed genes were related to metabolic pathways, neutrophil extracellular trap formation, alcoholism and transcriptional mis-regulation in cancer. A recently published study from China also showed that the mRNA expression profile change significantly at various stages of gallbladder cancer, and lipid-based metabolic abnormalities play an important role (Yang S et al., 2023). To discern their role in the context of *Salmonella* infection, several potential cancer associated candidates, which were significantly dysregulated, were further analysed. The data showed the importance of these genes in gallbladder tumorigenesis. One of the identified proteins ADAMTSL5, is a modulator of extracellular matrix. Interestingly, in our study the promoter region of ADAMTSL5 was found to be epigenetically modulated by KDM6B. KDM6B mediated remodeling of ADAMTSL5 promoter region was seen. In a recent work, ADAMTSL5 was reported to be associated with signalling of hepatocellular carcinoma, suggesting it to act as a master regulator of tumorigenicity (Arechederra et al., 2021).

Although we have shown that chronic *Salmonella* infection may contribute to GB carcinogenesis, there could be other bacteria or co-factors that might also play an important role, as suggested by the 'alpha-bug hypothesis' (Sears CL et al., 2011) or the 'driver–passenger model' (Tjalsma H et al., 2012) where a dominant microbe by itself may play a necessary but not sufficient role and could lead to other changes in the microenvironment that promote carcinogenesis. Although we could not fulfil Koch's postulate, it is difficult to apply Koch's postulate to cancers that are multifactorial in origin, and chronic infection could be one of the important factors.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Taken together, our findings suggest that there is a strong causal association between GBC and a particular type of chronic *Salmonella* Typhimurium infection (ST19), which probably originates from the gut, colonizes the gallbladder, induces chronic inflammation because of its virulence and causes epigenetic changes leading to gallbladder carcinogenesis. Our findings might explain the high incidence of GBC in northern India, where typhoid fever is endemic and the prevalence of gallbladder stones is ~5% in the general population.

Methods

The study was conducted after obtaining due clearance from the Institute Ethics Committee. Informed written consent was taken from all human participants.

Human Participants:

Group A: Patients with histology proven GBC were included as cases

Group B: Patients with gallbladder stone diseases were included as diseased controls

Inclusion criteria of cases and controls:

- Age 18-60 years
- No obstructive jaundice
- No prior surgical or endoscopic intervention

Exclusion criteria:

- Use of antibiotics in the preceding 4 weeks
- Use of Proton pump inhibitors in the preceding 4 weeks
- Metastatic disease
- Serum bilirubin >2 mg/dL
- Major comorbid illness, uncontrolled diabetes

Diagnosis of GBC: The diagnosis of GBC was suspected clinically and on imaging (contrast enhanced computed tomography scan of the abdomen), and confirmed on cytology or histopathology. The diagnosis of gallstone disease was made on an ultrasound examination of the abdomen.

Fecal and gallbladder tissue samples collection for microbiome study: Fecal sample was collected from the cases and controls after explaining the method of sample collection in a sterile container-containing sample transport buffer BiomLife (Ruhvenile Biomedical, India). Each participant was provided with a sterile plastic sheet and a sterile container with spoon. They were asked to get their hands and genital area cleaned with soap and water. In a dry place, the sterile

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

plastic sheet was opened just before defecation. Patient was advised to defecate at the center of the sheet without touching it. Then a scoop of fecal sample was taken with the spoon (approx. 3g) and transferred to the container-containing sample transport buffer BiomLife. The container was tightly closed and immediately transferred to 4°C before transporting to the laboratory and frozen at -80°C within 4 hrs. Like fecal samples, gallbladder tissues were also collected in sterile BiomLife transport buffer for metagenomic DNA extraction and microbiome study.

16S rRNA based fecal microbiome diversity and statistical analysis:

Gut and gallbladder tissue microbiome of study subjects was investigated by 16S rRNA based targeted metagenomic sequencing. The quality of NGS reads covering V3-V4 regions of 16S rRNA gene were evaluated and merged by fastp program by keeping Phred Quality Score(Q) ≥20, read length ≥50, and removing 10 bases from head and tail. The fastp program has also merging facility of paired reads. So, all quality passed paired reads were merged and used to generate operational taxonomic units (OTUs) by QIIME2-2021.11 pipeline (Bolyen et al., 2019). Two R version 3.6.1 packages, ggplot2 and ampViz2 were used to perform statistical analysis.

Quality assessment of NGS data

The quality of paired read generated from MiSeq platform was examined by FastQC version 0.11.9 program (Andrews, S. 2010). All paired reads were filtered by setup Phred Quality Score (Q) ≥20, read length ≥36, and removing 10 bases from the most sequencing error prone region i.e., head and tail of each read. These parameters were implemented in Trimmomatic program ver0.39 (Bolger et al, 2014) to get high quality reads.

Isolation, confirmation and cultivation of *Salmonella* Typhimurium from Gall bladder Tissue samples

Tissue samples from GBC patients were obtained during surgical resection of the tumor, which were confirmed by histology. The tumor tissue was collected in BHI medium and immediately transported to the lab at room temp. Collected tissues were enriched in 3 ml of BHI medium in aerobic conditions at 37º C. Following 6 hours incubation, 4 to 5 sterile glass beads (3.00 mm) were added into the medium and vortex gently for detaching cells from the tissue. Approximately, 300 µl culture aliquot was utilized for plating on the Heketon enteric agar plates. All the plates were incubated at 37°C for 48 hours. At least three colonies from each plate were confirmed by complete 16S rRNA gene sequencing from ten different GBC patients.

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Whole genome sequencing, quality assessment, assembling and annotation

Whole genome sequencing (WGS) was performed by a high-throughput next-generation Illumina MiSeq sequencing platform (*in-house*) and Nextera XT DNA Library preparation kit (Illumina, Inc., USA). High-quality raw reads (Q>30) were used for *de-novo* assembly by SPAdes v3.15.3 (Bankevich et al, 2012). The assembled genomes were annotated using Rapid Annotation using Subsystem Technology (RAST) version 2.0. The genomes were submitted in NCBI database. A global representation of *S.* Typhimurium deposited in Enterobase database has been included for comparative genomics with all the genomes downloaded from https://enterobase.warwick.ac.uk/species/index/senterica. At the time of this analysis, a total of 39565 *S.* Typhimurium genomes were available (as on 09.03.2022). For the present study, only genomes that had information for source, year and place of isolation sourced from humans have been considered. This yielded a total of 3665 genomes that fulfilled all the inclusion criteria, with STs belonging to eburst group 1 (eBG) ST19, ST34 and ST313. MLST STs of the present study isolates were determined using the pubMLST database schemes with seven housekeeping genes (*aroC, dnaN, hemD, hisD, purE, sucA* and *thrA*) (https://github.com/tseemann/mlst).

Phylogenetic analysis

The genome assemblies of *S.* Typhimurium were mapped against the reference *S.* Typhimurium LT2 (Accession No. NC_003197.2) using Snippy v4.6.0 (Seemann, 2015) (https://github.com/tseemann/snippy) (Seeman, 2015). The core genome SNP differences against the reference and as well within the genomes were generated as a core alignment file. SNPs were extracted from the core alignment https://github.com/sanger-pathogens/snp-sites, which was then subjected to construct a maximum likelihood (ML) phylogenetic analysis trough RaxML with GTRGAMMA model (Stamatakis et al., 2014). The generated phylogeny was then rooted against the reference genome *S.* Typhimurium LT2, visualized and annotated with the metadata using iTOL (Letunic and Bork, 2019). Among the 3665 genomes included in the preliminary analysis, genomes that were identical and sourced from an outbreak clusters were removed. The final genomes that were included in the further analysis were 2517, with 28 and 2489 genomces from GBC and global representatives respectively.

Mutational profiling

The presence of any type of nucleotide variation as mutations in the form of single nucleotide polymorphisms (SNPs), insertions or deletions was identified *in-silico* using snippy pipeline with mapping and variant calling features. To obtain the mutational profile, the genomes

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

assemblies were compared against annotated reference genome *S.* Typhimurium LT2 (Accession No. NC_003197.2). Mutation accumulation was manually curated with the respective STs included in the study. Importantly, genes that were inactivated by means of either missense mutations or an insertion/deletion leading to frameshift mutations were analyzed. Any such mutations that resulted in the loss of functionality of the protein were termed as hypothetically disrupted coding sequences (HDCS) or pseudogenes. These genes were compared with the HDCS host adaptation signatures of the invasive ST313 lineages.

Pan genome analysis

The pan genome analysis was performed for all the 2517 genomes and reference LT2. The genomes were annotated using using Prokka v. 1.14 (Seemann et al., 2014), with the genus database "*Salmonella*" (https://github.com/tseemann/prokka). The GFF files that were generated from prokka was taken to assess the genome diversity of the 2517 genomes using panaroo (Tonkin-Hill et al., 2020) (https://github.com/gtonkinhill/panaroo) (Tonkin-Hill et al., 2020). Panaroo was run with "strict" mode option that enables to retain genes that were present at least in 5% of all the genomes being analysed. The gene presence/absence results obtained from panaroo was run through twilight scripts and grouped as GBC invasive ST19 (n=28), global sources: ST19 (n=1880), ST34 (n=418) and ST313 (n=191) (https://github.com/ghoresh11/twilight) (Horesh et al., 2021)

***In-silico* identification of plasmid and prophage sequences**

PLSDB was used to check for the plasmid and plasmid replicons within the studied genome of *S.* Typhimurium, n=28 (Galata et al., 2019). The analysis was completed using Mash (search strategy: mash screen) with a maximal p-value of 0.1 and minimal identity of 0.99. To remove redundancy from the output data, the winner-takes-all strategy was applied. 28 genomes from *S.* Typhimurium were analyzed for the presence of prophage sequences using PHASTER (PHAge Search Tool Enhanced Release) (http://phaster.ca/) web server (Arndt et al., 2016). It is a upgradation of the PHAST webserver for the fast identification of prophage sequences present within the bacterial genome (Zhou et al., 2011). Both the plasmid and prophage sequences were further confirmed for its presence in the studied *S.* Typhimurium genomes through BLASTn manually.

Detection of virulence factors in the genome of *S.* Typhimurium

Virulence factor database (VFDB) was used for the detection of putative virulence factors in the *S.* Typhimurium genome (n=28) isolated from GBC tissue samples (**Liu et al., 2022**). VF analyser tool available at VFDB are developed in Java-Script rich manner. The work flow proceeds by construction of orthologous groups within the query genome and the previously

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

analysed *S.* Typhimurium strain LT2 as the representative genome to get rid of false positive results. Next it carries out repetitive similarity searches for the accurate detection of virulence factors (VF). Eventually a data refinement process is carried out for VF encoded by gene clusters (Liu et al., 2022). We also manually carried out BLASTn searches for the proinflammatory and anti-inflammatory virulence genes present in the reference *S.* Typhimurium *LT2* genome with our draft genome assemblies and compared it with the results of VF analyser. The virulence genes encoding virulence factors for the reference strain LT2 were downloaded from VFDF dataset. Heatmap representing virulence genes were prepared in R software.

Genome-wide expression profiling of human GBC:

Tissue samples were collected from GBC and gallstone patients during surgery. Each sample was divided and put into two tubes: one containing RNA later and the other containing formalin. Samples in RNA later were initially stored at 4˚ C for 24 hours and then stored at -80° C.

Tissue samples in formalin were processed for block formation. H&E staining was done to confirm the presence of tumor area.

Total-RNA extraction from tissue was standardized using organic method (Ambion extraction kit).

Controls: We included histologically normal mucosa from patients with gallbladder stones as control.

RNA extraction and quality check

RNA extraction was done using mirVana microRNA isolation kit (Ambion). The quality of extracted RNA was checked by nanodrop and 260/280 value ranged between 1.8 to 2.0. Quality and integrity of extracted RNA was also checked using Agilent 2100 Bioanalyzer, RNA 6000 Nano LabChip kit and Agilent 2100 Expert Software (Agilent Technologies, Santa Clara, CA, USA). Samples with RNA integrity number (RIN) more than or equal to 6.5 were selected for further analysis.

Microarray experiment

The protocol of one-color micro-array-based gene expression (Agilent technologies; SurePrint G3 Human Gene Expression 8x60K v2 Microarray) was followed as per manufacturer's recommendations. Hybridized arrays were scanned with Agilent's G2505B microarray scanner system and the raw data was collected. Out of 35 samples collected, the microarray dataset consisted of 32 samples, of which 16 samples were control and 16 were from adenocarcinoma patients.

Differential gene expression analysis

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

The microarray data generated from Agilent platform belonged to two categories adenocarcinoma (adeno) and the control group. The LIMMA (Bioconductor; https://www.bioconductor.org/) package has been used to convert the probe level data into the expression measures and DEGs were analyzed in the adeno vs control group. DEGs were reported using the criteria of fold-change ≥1.5 and fold change < −1.5 by conducting t-test with adjusted p value of <0.05, characterized as upregulated and downregulated genes. Benjamini Hochberg correction was applied.

For functional analysis, KEGG (Kyoto Encyclopaedia of Genes and Genomes) (http://www.kegg.jp/) analysis was performed on differentially expressed genes to understand the pathways involved.

Experimental murine studies:

A. **Preparation of *S.* Typhimurium bacterial culture for infection**: Fresh *S.* Typhimurium (SL1344 and SL3261 are laboratory strain) was revived from -80^0 C storage and streaked on a fresh Luria Bertani (LB) Agar plate. On the day before experiment, a single colony was picked from the LB agar plate and inoculated to a 2 ml of Luria broth (LB) taken in a sterile falcon tube and grown at 37^0 C with aeration. For secondary culture, grown primary culture was added to 12 ml of LB at 1:1000 dilution and was grown over night at 37^0 C without shaking. On the day of infection, cells were pelleted at 4000 rpm for 10 min and washed with filtered 1X PBS for two times and resuspended in 1 ml 1X PBS. The bacteria were then added to tissue culture plate containing mammalian cells at MOI of 1:40 for NOZ cells and MOI of 1:10 for primary gallbladder cells isolated from mice gallbladder.

B. **Cell and culture conditions:** Human gallbladder carcinoma cell line NOZ (procured from **JCRB** Cell Bank - Japanese Collection of Research Bioresources Cell Bank) were grown in DMEM media (Sigma) supplemented with 14 mM NaHCO3 (Sigma), 15 mM HEPES buffer (pH 7.4) (GIBCO), 1 mM sodium pyruvate (GIBCO), 40 mg/ L penicillin (GIBCO), 90 mg/L streptomycin (GIBCO), and 10% fetal bovine serum (GIBCO). Gallbladder primary cell line were prepared from C57BL/6 as previously described (Rahul Kuver; 1996). All cell cultures were incubated in humidified 37°C incubator with 5% CO_2.

C. ***Salmonella* Typhimurium infection of cultured mammalian cells**: Cells were grown and seeded in tissue culture treated appropriate well plates. Upon attaining confluency, cells were washed and again supplemented with antibiotic free media. NOZ Cells were then infected with *S.* Typhimurium SL1344 as mentioned above at MOI 1:40 and gallbladder primary cells at MOI of 1:10. This setup was incubated at 37ºC for 1 h. Following incubation, bacteria were washed off two times with media and resuspended again in fresh media containing 100ug/ml gentamicin for 1 h at 37ºC. After 1hr, media was again removed and

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

fresh media with 10ug/ml gentamicin was supplemented. Cells were then incubated for the rest of period of infection at 37^0 C in CO_2incubator. The cells were then harvested for protein and RNA which were further analysed by western blot and real time PCR.

D. ***Salmonella in vivo*** **chronic carriage gallstone model**: The FVB/N mice of 3-4 weeks were taken and grouped into Control, SL infected, Lithogenic diet (LD) and Lithogenic diet plus*Salmonella* infected (LD-SL). LD and LD-SL grouped mice were fed for 9 weeks a lithogenic diet i.e. 1% cholesterol and 0.5% cholic acid mixed with the normal diet. Post 9 weeks, the chronic infection causing bacteria SL3261ΔaroA was fed orally to mice at 109 CFU by gavage. Mice were euthanized at different time intervals (15 days, 4 months, 8.5 months, 1.5 years and 1.8 years) post infection and gallbladder was taken out and processed for histopathogical study, RNA, and protein. For subcutaneous tumor mice model, GSKJ4 drug treatment was given at a dose of 1mg/Kg of mice body weight intraperitoneally (ip) on alternative days for 1 month. Appropriate drug control and untreated control were included in the study.

E. **Immunoblotting**: Protein samples were prepared from respective mice tissue and mentioned cell lines by lysis with RIPA lysis buffer (Sigma R0278) and 1X protease inhibitor cocktail (G-Biosciences). The cell lysates prepared were quantified using CBX protein assay kit (G-Bioscience, USA). Equal amount of protein was resolved on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and then transferred to nitrocellulose membrane. Blots were probed with appropriate antibodies.

F. **RNA extraction and quantitative real-time PCR (qRT-PCR)**: Total RNA were isolated from mice tissues, human GBC cell line NOZ and organoids using Nucleo Spin RNA-II Kit (MN, Germany) according to the manufacturer's protocol. One microgram of total RNA from each sample was used to synthesize cDNA using i-Script cDNA synthesis Kit (Bio-Rad, USA). Bio-Rad CFX 96TM Real-Time Detection System was used for real-time PCR (qRT-PCR) using i-Taq Syber green (BioRad, USA) according to manufacturer's instruction. Gene expression levels were normalized according to the average cycle threshold values for the internal control gene HPRT. Cycle threshold values were extracted and data analysis was performed using the ΔΔCt method.

G. **Histopathology and Immunohistochemistry:** In order to perform histopathology and immunohistochemistry of gallbladder tissue, sections of tissue were fixed in 10% formalin overnight and embedded in parafilm. Sections of 5 mm thickness were cut onto glass slides and processed using Cryotome (Thermo) and were stained with hematoxylin and eosin (H&E) and evaluated under a light microscope with attached DP25 digital camera. For IHC,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

the tissue glass slides were washed using 1X PBS for three times (5 minutes each). Further, the endogenous peroxidase activity was quenched by treating the tissue with 3% H_2O_2 for 20 minutes. The tissue sample or cells were again washed using 1X PBS for two times (5 minutes each), then blocked with 5% goat serum at room temperature for one hour. The sections were incubated with primary antibodies prepared in 5% goat serum overnight in a moist chamber. Sections were washed with 1X PBS for three times at 5 minutes each and were incubated with secondary HRP conjugate antibody for 2 hours prepared in 5% goat serum at room temperature in a moist chamber. The tissues were washed three times using 1X PBS. For immunohistochemistry, the slides were stained with DAB substrate (Sigma, USA) and counterstained with Haematoxylin. Images were taken on a Nikon Fluorescence Microscope using colour camera.

H. **Cell migration using wound healing assay**: The KDM6B gene knockdown and knocked down effect on cell migration was examined using wound healing assay. NOZ cells were plated at a density of 1.5×10^5/well in DMEM with 1% FBS in a 24-well plate and cultured. After approximately 24 hours of incubation, the cells were transfected with KDM6B WT and catalytic deficient plasmid. Next, the monolayer were scraped in a straight line to create a 'scratch' with a p200 pipette tip as reported earlier (Mustafa *et. al.,* 2017). Plates were then incubated and monitored periodically by microscopy to examine wound and migration of cells in transfected and un-transfected cells.

I. **Cell proliferation assay**: NOZ cells were seeded (1.5×10^5 cells /well) onto the cell culture plates. After approximately 24 hours of incubation, the cells were transfected with KDM6B WT and catalytic deficient plasmid. Next, these cells were split on the 6 well plates after counting equal no. of cells i.e. 1000 per well. After 1 week, the bottom surface of the plate was stained with 0.1% crystal violet and quantified using a light microscope.

J. **ChIP qPCR expression analysis**:

NOZ cells were grown to confluency of 80-90% and infected with *Salmonella* for 18 hrs as described previously. Post 18 hrs of infection, the cells were cross linked with 1% formaldehyde at RT for 10 mins. Crosslinking was stopped by addition of 125mM glycine to quench formaldehyde. The cells were collected and washed with PBS 3 times. Cells were then lysed with lysis buffer (PBS + 0.1% NP40 + protease inhibitor) to obtain nucleus. The nuclei were lysed using nucleus lysis buffer for 10 mins on ice. The lysed samples were sheared using a sonicator to obtain chromatin fragments of 200-500 bp. Sheared chromatin was subjected to immunoprecipitation using ChIP grade antibodies specific for KDM6B, Histone3 K27me3 and IgG control acting as negative control overnight at 4^0C. The immunoprecipitated DNA was pulled using Protein G sepharose beads through incubation at

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

4^0C for 3 hrs. The immunoprecipitated DNA and the input DNA were purified. qRT-PCR was performed using shortlisted promoter sequences of wnt pathway genes.

K. PCR identification for *Salmonella* isolates

To detect the persistence of *Salmonella* in the mice, different organs were isolated and bacterial burden assay was performed in the above- described mouse model system. Mice were euthanized post infection and gallbladder and liver were harvested and homogenized in PBS (.1% triton X-100). The homogenized tissues were then serially diluted and plated onto Brilliant green agar and Hekton agar plates. The plates were incubated in a 37-degree incubator and bacterial CFU counted. All animal experiments were carried out in the Animal Facility of Regional Center for Biotechnology (RCB). Animal ethics proposal were approved by the RCB Institutional Animal Ethics Committee (approval no. IAEC/RCB/2021/102 and IAEC/RCB/ 2019/062).

L. Statistical Analysis:

All results are expressed as the mean standard error from an individual experiment done in triplicate. Data were analyzed with one way ANOVA followed by standard unpaired two-tailed Student's t-test, Tukey's post-test and the Mann–Whitney U-test as applicable, with p-values of <0.05–0.001 were considered statistically significant. We used GraphPad PRISM for statistical analysis.

Acknowledgement: This research is supported by Intramural research grants by All India Institute of Medical Sciences, New Delhi to Pramod Kumar Garg, Kajal Jain, Asuri Krishna and VK Bansal. Dr. Pramod Kumar Garg is supported by the JC Bose fellowship by the Science & Engineering Research Board. Ethics clearances from the following IBSC, IAEC and IEC were obtained: (IEC-495/02.08.2019, RCB/IAEC/2019/062, RCB/IAEC/2021/102, RCB/IBSC/21-22/341, RCB-BBB-IEC-HI-31, RCB/IBSC/22-23/405, RCB/IAEC/2023/158, RCB/IBSC/23-24/480). We also thank the instrumentation facility at RCB and Small animal facility at NCR-Biotech Science Cluster. BD acknowledges Department of Biotechnology, Govt. of India for financial support (No. BT/PR30159/MED/15/188/2018). Authors acknowledge Mr. Subhash Tanwar, Mr. Roshan Kumar and Advanced Nucleotide Sequencing Facility of THSTI for microbial genome sequencing.

Author contributions: PKG conceived the idea; PKG, CVS, BD, KJ designed the experiments, supervised research, arranged resources and wrote the paper; AKP, SM, KJ, SP,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

CR, SK, AP, DP, KK, DT, ANS, JV, PJ, SR, PK, AK, RY, BD performed experiments and contributed in the preparation of manuscript draft. SK, VKB provided clinical resources, designed experiments and helped in data analysis. All the authors approved the final manuscript.

Declaration of interests: Authors declared that there are no financial or other competing interests.

Figure legends:

Figure 1: Boxplot is displaying relative abundance of 10 bacterial phyla. The y- and x-axis represent relative abundance and phylum name, respectively. The statistical significance value was calculated by ANOVA for each phylum among three groups. The significant value (p-value) viz. 'ns' denotes non-siginifcant. '*', '**', '***' and '****' represent the level of significant as ≤ 0.05, ≤0.001, ≤0.0001, and ≤0.00001, respectively.

Figure 2: (a) Maximum likelihood phylogenetic analysis of *S.* Typhimurium isolated from gallbladder tissues of GBC patients and compared with the global collection of 2490 genomes. Scale bar indicates number of substitution per genome/per site. **(b)** sub-tree showing the high level resolution of the sub-cluster of ST19 genomes of GBC with other genomes from the global collection

Figure 3: (a) Dot plot shows the invasive index score of each of the *S.* Typhimurium isolates that was determined against the non-invasive reference *S.* Typhimurium genome LT2. **(b)** shows the distribution chart for the mutation frequency categorized as frameshift, missense, synonymous and others.

Figure 4: Pan genome analysis plots of 2518 *S.* Typhimurium genomes. **(a)** Phandango visualization of phylogenetic tree with the presence/absence gene profile, **(b)** sub-classification of the gene categories with respect to the distribution frequency, **(c)** mean distribution frequency as classified with different STs, **(d)** number of gene distributed with different STs groups.

Figure 5. Development of murine model of gallstone and chronic *Salmonella* infection [A] Schematic representation of the steps involved in development of Gallstone mice along with chronic Salmonella infection model Post Salmonella (ST) infection at the indicated time, the mice were euthanized followed by isolation of gallbladder, spleen and liver. Labels: C for mock-infected and normal diet fed control group, ST for Salmonella infection group, LD for Lithogenic diet (involving 1% cholesterol and 0.5% Cholic acid) group, LD-ST for

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Lithogenic diet and Salmonella infection group. [B] Gallbladders from the indicated group along with magnified view of the generated gallstones (inset) are represented. [C] Splenomegaly seen in infected mice along with enlarged liver seen with change in color in Gallstones. [D] Agarose gel electrophoresis image post PCR amplification of DNA from isolated bacterial colonies cultured from infected mice (15 days post infection) from liver tissue using *Salmonella* Typhimurium 16S specific primers. [E] Agarose gel images of similar experiment carried out as previous figure with Fecal pellets to record shedding of *Salmonella*. DNA of *Salmonella* Typhimurium strain SL1344 was used as a positive control (+) and non DNA template was used as a negative control (-).

Figure 6. Detection of Histopathological alterations in Gallstone mice model. Representative images of Hematoxylin and Eosin stained gallbladder sections from the indicated groups. The infection with *Salmonella* was for either 4 months [A.] or 8.5 months [B.] The epithelial lining of gallbladder is marked by blue arrowheads, inflammation by arrowheads, fibrosis by black arrowheads, metaplasia by green arrowheads, and hyperplasia by arrowheads. The values of each of these parameters were quantified as histopathological score by a pathologist blinded to the group and plotted (Right panel). [C.] Periodic acid Schiff (PAS) staining of representative sections of the gallbladder of all the groups 4 months post Salmonella infection indicating presence of higher amounts of acidic mucins in LD-ST mice.

Figure 7. Alteration of epigenetic modulator KDM6B involved in maintenance of H3K27me3 mark upon *Salmonella* Typhimurium infection. Representation of KDM6B mRNA upregulation 15 days post ST infection, HPRT used for normalization of the data [A.] Immunoblot representing KDM6B upregulation in gallbladder tissue of different indicated groups of mice post infection of indicated duration. Expression analysis of P16 [C.] and P53[D.] [B.] Immunoblot representing increased Kdm6B expression in gallbladder cancer patient tissue (labelled with GBC1-3) compared to those from non-malignant control samples (labelled as NMC1-3) and gallstone control samples (labelled as GS1-3). Expression analysis of various genes [F.] Chromatin immunoprecipitation assay carried out using Kdm6B [G.], H3K27me3 [H.] and [I.] histone H3 antibodies for binding in ADAMTSL5 gene promoter post 18 hrs of Salmonella Typhimurium (SL1344) infection. Data representing fold enrichment was plotted in each case. Expression analysis of ADAMTSL5 in murine gallbladder post 15 days of *Salmonella* Typhimurium (SL1344 strain) infection represented in fold change values by Real Time PCR. HPRT was used for normalization [J.].

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Figure 8: (a) Photograph showing cut-open gallbladder (GB) with liver margin (L) having multiple growth/tumor (Tu) arising from fundus and two calculi (Ca). The region around the calculi appears to be free of tumor (PN). **(b)** Representative hematoxylin & eosin staining of a representative sample of gallbladder cancer **(c)** Post-extraction, total-RNA samples run on 1.5% agarose gel and bio-analyzer electropherogram of representative RNA sample **(d)** Volcano plot showing differentially expressed transcripts (480; 327 up-regulated and 153 down-regulated) in GBC patient samples compared to gallstone controls. **(e)** Functional analysis through KEGG pathway: KEGG enrichment analysis of differentially expressed genes (DEGs) between GBC and gallbladder with stones. X-axis represent the number of genes. The p-values for all the genes were significant but the corrected p-value was not found to be significant.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

References

1. Arechederra, M., Bazai, S. K., Abdouni, A., Sequera, C., Mead, T. J., Richelme, S., Daian, F., Audebert, S., Dono, R., Lozano, A., Gregoire, D., Hibner, U., Allende, D. S., Apte, S. S., & Maina, F. (2021). ADAMTSL5 is an epigenetically activated gene underlying tumorigenesis and drug resistance in hepatocellular carcinoma. *Journal of Hepatology*, *74*(4), 893–906. https://doi.org/10.1016/j.jhep.2020.11.008

2. Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–W21. https://doi.org/10.1093/nar/gkw387

3. Attili, A. F., De Santis, A., Capri, R., Repice, A. M., Maselli, S., & Group, G. (1995). The natural history of gallstones: The GREPCO experience. *Hepatology*, *21*(3), 656–660. https://doi.org/10.1002/hep.1840210309

4. Aykut, B., Pushalkar, S., Chen, R., Li, Q., Abengozar, R., Kim, J. I., Shadaloey, S. A., Wu, D., Preiss, P., Verma, N., Guo, Y., Saxena, A., Vardhan, M., Diskin, B., Wang, W., Leinwand, J., Kurz, E., Kochen Rossi, J. A., Hundeyin, M., … Miller, G. (2019). The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. *Nature*, *574*(7777), 264–267. https://doi.org/10.1038/s41586-019-1608-2

5. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

6. Batra, Y., Pal, S., Dutta, U., Desai, P., Garg, P. K., Makharia, G., Ahuja, V., Pande, G. K., Sahni, P., Chattopadhyay, T. K., & Tandon, R. K. (2005). Gallbladder cancer in India: A dismal picture. *Journal of Gastroenterology and Hepatology*, *20*(2), 309–314. https://doi.org/10.1111/j.1440-1746.2005.03576.x

7. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

8. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

9. Bruno, V. M., Hannemann, S., Lara-Tejero, M., Flavell, R. A., Kleinstein, S. H., & Galán, J. E. (2009). Salmonella Typhimurium Type III Secretion Effectors Stimulate Innate Immune Responses in Cultured Epithelial Cells. *PLoS Pathogens*, *5*(8), e1000538. https://doi.org/10.1371/journal.ppat.1000538

10. Carden, S. E., Walker, G. T., Honeycutt, J., Lugo, K., Pham, T., Jacobson, A., Bouley, D., Idoyaga, J., Tsolis, R. M., & Monack, D. (2017). Pseudogenization of the

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Secreted Effector Gene sseI Confers Rapid Systemic Dissemination of S. Typhimurium ST313 within Migratory Dendritic Cells. *Cell Host & Microbe*, *21*(2), 182–194. https://doi.org/10.1016/j.chom.2017.01.009

11. Chang, M.-H. (2014). Prevention of hepatitis B virus infection and liver cancer. *Recent Results in Cancer Research. Fortschritte Der Krebsforschung. Progres Dans Les Recherches Sur Le Cancer*, *193*, 75–95. https://doi.org/10.1007/978-3-642-38965-8_5

12. Chen, M., Zou, S., He, C., Zhou, J., Li, S., Shen, M., Cheng, R., Wang, D., Zou, T., Yan, X., Huang, Y., & Shen, J. (2020). Transactivation of SOX5 by Brachyury promotes breast cancer bone metastasis. *Carcinogenesis*, *41*(5), 551–560. https://doi.org/10.1093/carcin/bgz142

13. Crawford, R. W., Rosales-Reyes, R., Ramírez-Aguilar, M. D. L. L., Chapa-Azuela, O., Alpuche-Aranda, C., & Gunn, J. S. (2010). Gallstones play a significant role in *Salmonella* spp. Gallbladder colonization and carriage. *Proceedings of the National Academy of Sciences*, *107*(9), 4353–4358. https://doi.org/10.1073/pnas.1000862107

14. Dapito DH, Mencin A, Gwak GY, et al. Promotion of hepatocellular carcinoma by the intestinal microbiota and TLR4. Cancer Cell 2012; 21:504–16.

15. De Groote, M. A., Ochsner, U. A., Shiloh, M. U., Nathan, C., McCord, J. M., Dinauer, M. C., Libby, S. J., Vazquez-Torres, A., Xu, Y., & Fang, F. C. (1997). Periplasmic superoxide dismutase protects *Salmonella* from products of phagocyte NADPH-oxidase and nitric oxide synthase. *Proceedings of the National Academy of Sciences*, *94*(25), 13997–14001. https://doi.org/10.1073/pnas.94.25.13997

16. Di Domenico, E. G., Cavallo, I., Pontone, M., Toma, L., & Ensoli, F. (2017). Biofilm Producing Salmonella Typhi: Chronic Colonization and Development of Gallbladder Cancer. *International Journal of Molecular Sciences*, *18*(9), 1887. https://doi.org/10.3390/ijms18091887

17. Ding, S.-Z., Goldberg, J. B., & Hatakeyama, M. (2010). *Helicobacter pylori* infection, oncogenic pathways and epigenetic mechanisms in gastric carcinogenesis. *Future Oncology*, *6*(5), 851–862. https://doi.org/10.2217/fon.10.37

18. Dutta, U., Garg, P. K., Kumar, R., & Tandon, R. K. (2000). Typhoid carriers among patients with gallstones are at increased risk for carcinoma of the gallbladder. *The American Journal of Gastroenterology*, *95*(3), 784–787. https://doi.org/10.1111/j.1572-0241.2000.01860.x

19. Dutta, U., Nagi, B., Garg, P. K., Sinha, S. K., Singh, K., & Tandon, R. K. (2005). Patients with gallstones develop gallbladder cancer at an earlier age: *European Journal of Cancer Prevention*, *14*(4), 381–385. https://doi.org/10.1097/00008469-200508000-00011

20. Ethun, C. G., Le, N., Lopez-Aguiar, A. G., Pawlik, T. M., Poultsides, G., Tran, T., Idrees, K., Isom, C. A., Fields, R. C., Krasnick, B. A., Weber, S. M., Salem, A., Martin, R. C. G., Scoggins, C. R., Shen, P., Mogal, H. D., Schmidt, C., Beal, E., Hatzaras, I., … Maithel, S. K. (2017). Pathologic and Prognostic Implications of

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Incidental versus Nonincidental Gallbladder Cancer: A 10-Institution Study from the United States Extrahepatic Biliary Malignancy Consortium. *The American Surgeon*, *83*(7), 679–686. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5915617/

21. Everhart, J. E., & Ruhl, C. E. (2009). Burden of Digestive Diseases in the United States Part III: Liver, Biliary Tract, and Pancreas. *Gastroenterology*, *136*(4), 1134–1144. https://doi.org/10.1053/j.gastro.2009.02.038

22. Farrant, J. L., Sansone, A., Canvin, J. R., Pallen, M. J., Langford, P. R., Wallis, T. S., Dougan, G., & Kroll, J. S. (1997). Bacterial copper- and zinc-cofactored superoxide dismutase contributes to the pathogenesis of systemic salmonellosis. *Molecular Microbiology*, *25*(4), 785–796. https://doi.org/10.1046/j.1365-2958.1997.5151877.x

23. Feasey, N. A., Masesa, C., Jassi, C., Faragher, E. B., Mallewa, J., Mallewa, M., MacLennan, C. A., Msefula, C., Heyderman, R. S., & Gordon, M. A. (2015). Three Epidemics of Invasive Multidrug-Resistant *Salmonella* Bloodstream Infection in Blantyre, Malawi, 1998–2014. *Clinical Infectious Diseases*, *61*(suppl 4), S363–S371. https://doi.org/10.1093/cid/civ691

24. Figueroa-Bossi, N., & Bossi, L. (1999). Inducible prophages contribute to Salmonella virulence in mice. *Molecular Microbiology*, *33*(1), 167–176. https://doi.org/10.1046/j.1365-2958.1999.01461.x

25. Figueroa-Bossi, N., Uzzau, S., Maloriol, D., & Bossi, L. (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. *Molecular Microbiology*, *39*(2), 260–272. https://doi.org/10.1046/j.1365-2958.2001.02234.x

26. Francescone R, Hou V, Microbiome GSI. inflammation, and cancer. Cancer J 2014; 20:181–9.

27. Galata, V., Fehlmann, T., Backes, C., & Keller, A. (2019). PLSDB: A resource of complete bacterial plasmids. *Nucleic Acids Research*, *47*(D1), D195–D202. https://doi.org/10.1093/nar/gky1050

28. Gulig, P. A., Danbara, H., Guiney, D. G., Lax, A. J., Norel, F., & Rhen, M. (1993). Molecular analysis of spv virulence genes of the salmonella virulence plasmids. *Molecular Microbiology*, *7*(6), 825–830. https://doi.org/10.1111/j.1365-2958.1993.tb01172.x

29. Heberle, H., Meirelles, G. V., Da Silva, F. R., Telles, G. P., & Minghim, R. (2015). InteractiVenn: A web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, *16*(1), 169. https://doi.org/10.1186/s12859-015-0611-3

30. Heffernan, E. J., Reed, S., Hackett, J., Fierer, J., Roudier, C., & Guiney, D. (1992). Mechanism of resistance to complement-mediated killing of bacteria encoded by the Salmonella typhimurium virulence plasmid gene rck. *Journal of Clinical Investigation*, *90*(3), 953–964. https://doi.org/10.1172/JCI115972

31. Helmuth, R., Stephan, R., Bunge, C., Hoog, B., Steinbeck, A., & Bulling, E. (1985). Epidemiology of virulence-associated plasmids and outer membrane protein patterns

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

within seven common Salmonella serotypes. *Infection and Immunity*, *48*(1), 175–182. https://doi.org/10.1128/iai.48.1.175-182.1985

32. Hiley, L., Graham, R. M. A., & Jennison, A. V. (2019). Genetic characterisation of variants of the virulence plasmid, pSLT, in Salmonella enterica serovar Typhimurium provides evidence of a variety of evolutionary directions consistent with vertical rather than horizontal transmission. *PLOS ONE*, *14*(4), e0215207. https://doi.org/10.1371/journal.pone.0215207

33. Horesh, G., Taylor-Brown, A., McGimpsey, S., Lassalle, F., Corander, J., Heinz, E., & Thomson, N. R. (2021). Different evolutionary trends form the twilight zone of the bacterial pan-genome. *Microbial Genomics*, *7*(9). https://doi.org/10.1099/mgen.0.000670

34. Hulsegge, I., Kommadath, A., & Smits, M. A. (2009). Globaltest and GOEAST: Two different approaches for Gene Ontology analysis. *BMC Proceedings*, *3*(S4), S10. https://doi.org/10.1186/1753-6561-3-S4-S10

35. Iyer, P., Barreto, S. G., Sahoo, B., Chandrani, P., Ramadwar, M. R., Shrikhande, S. V., & Dutt, A. (2016). Non-typhoidal Salmonella DNA traces in gallbladder cancer. *Infectious Agents and Cancer*, *11*(1), 12. https://doi.org/10.1186/s13027-016-0057-x

36. Jacob, J. J., Anandan, S., Venkatesan, M., Neeravi, A., Vasudevan, K., Pragasam, A. K., & Veeraraghavan, B. (2019). Genomic analysis of human invasive Salmonella enterica serovar Typhimurium ST313 isolate B3589 from India. *Infection, Genetics and Evolution*, *73*, 416–424. https://doi.org/10.1016/j.meegid.2019.05.023

37. Jacob, J. J., Solaimalai, D., Muthuirulandi Sethuvel, D. P., Rachel, T., Jeslin, P., Anandan, S., & Veeraraghavan, B. (2020). A nineteen-year report of serotype and antimicrobial susceptibility of enteric non-typhoidal Salmonella from humans in Southern India: Changing facades of taxonomy and resistance trend. *Gut Pathogens*, *12*(1), 49. https://doi.org/10.1186/s13099-020-00388-z

38. Jain, K., Mohapatra, T., Das, P., Misra, M. C., Gupta, S. D., Ghosh, M., Kabra, M., Bansal, V. K., Kumar, S., Sreenivas, V., & Garg, P. K. (2014). Sequential Occurrence of Preneoplastic Lesions and Accumulation of Loss of Heterozygosity in Patients With Gallbladder Stones Suggest Causal Association With Gallbladder Cancer. *Annals of Surgery*, *260*(6), 1073–1080. https://doi.org/10.1097/SLA.0000000000000495

39. Jain, K., Sreenivas, V., Velpandian, T., Kapil, U., & Garg, P. K. (2013). Risk factors for gallbladder cancer: A case-control study. *International Journal of Cancer*, *132*(7), 1660–1666. https://doi.org/10.1002/ijc.27777

40. Kingsley, R. A., Msefula, C. L., Thomson, N. R., Kariuki, S., Holt, K. E., Gordon, M. A., Harris, D., Clarke, L., Whitehead, S., Sangal, V., Marsh, K., Achtman, M., Molyneux, M. E., Cormican, M., Parkhill, J., MacLennan, C. A., Heyderman, R. S., & Dougan, G. (2009). Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Research*, *19*(12), 2279–2287. https://doi.org/10.1101/gr.091017.109

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

41. Koshiol, J., Wozniak, A., Cook, P., Adaniel, C., Acevedo, J., Azócar, L., Hsing, A. W., Roa, J. C., Pasetti, M. F., Miquel, J. F., Levine, M. M., Ferreccio, C., the Gallbladder Cancer Chile Working Group, Aguayo, C. G., Baez, S., Díaz, A., Molina, H., Miranda, C., Castillo, C., … Hildesheim, A. (2016). *Salmonella enterica* serovar Typhi and gallbladder cancer: A case–control study and meta-analysis. *Cancer Medicine*, *5*(11), 3310–3235. https://doi.org/10.1002/cam4.915

42. Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, *47*(W1), W256–W259. https://doi.org/10.1093/nar/gkz239

43. Liu, B., Zheng, D., Zhou, S., Chen, L., & Yang, J. (2022). VFDB 2022: A general classification scheme for bacterial virulence factors. *Nucleic Acids Research*, *50*(D1), D912–D917. https://doi.org/10.1093/nar/gkab1107

44. Lou, L., Zhang, P., Piao, R., & Wang, Y. (2019). Salmonella Pathogenicity Island 1 (SPI-1) and Its Complex Regulatory Network. *Frontiers in Cellular and Infection Microbiology*, *9*, 270. https://doi.org/10.3389/fcimb.2019.00270

45. Mambu, J., Barilleau, E., Fragnet-Trapp, L., Le Vern, Y., Olivier, M., Sadrin, G., Grépinet, O., Taieb, F., Velge, P., & Wiedemann, A. (2020). Rck of Salmonella Typhimurium Delays the Host Cell Cycle to Facilitate Bacterial Invasion. *Frontiers in Cellular and Infection Microbiology*, *10*, 586934. https://doi.org/10.3389/fcimb.2020.586934

46. Marchello, C. S., Dale, A. P., Pisharody, S., Rubach, M. P., & Crump, J. A. (2019). A Systematic Review and Meta-analysis of the Prevalence of Community-Onset Bloodstream Infections among Hospitalized Patients in Africa and Asia. *Antimicrobial Agents and Chemotherapy*, *64*(1), e01974-19. https://doi.org/10.1128/AAC.01974-19

47. Maringhini, A. (1987). Gallstones, Gallbladder Cancer, and Other Gastrointestinal Malignancies: An Epidemiologic Study in Rochester, Minnesota. *Annals of Internal Medicine*, *107*(1), 30. https://doi.org/10.7326/0003-4819-107-1-30

48. Dudgaard, A., & Gaarslev, K. (1988). Risk of Hepatobiliary Cancer in Carriers of Salmonella typhi. *JNCI Journal of the National Cancer Institute*, *80*(4), 288–288. https://doi.org/10.1093/jnci/80.4.288

49. Nath, G., Singh, Y. K., Kumar, K., Gulati, A. K., Shukla, V. K., Khanna, A. K., Tripathi, S. K., Jain, A. K., Kumar, M., & Singh, T. B. (2008). Association of carcinoma of the gallbladder with typhoid carriage in a typhoid endemic area using nested PCR. *The Journal of Infection in Developing Countries*, *2*(04), 302–307. https://doi.org/10.3855/jidc.226

50. N esić D, Hsu Y, Stebbins CE. Assembly and function of a bacterial genotoxin. Nature 2004; 429:429–33.

51. Okoro, C. K., Barquist, L., Connor, T. R., Harris, S. R., Clare, S., Stevens, M. P., Arends, M. J., Hale, C., Kane, L., Pickard, D. J., Hill, J., Harcourt, K., Parkhill, J., Dougan, G., & Kingsley, R. A. (2015). Signatures of Adaptation in Human Invasive

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Salmonella Typhimurium ST313 Populations from Sub-Saharan Africa. *PLOS Neglected Tropical Diseases*, *9*(3), e0003611. https://doi.org/10.1371/journal.pntd.0003611

52. Okoro, C. K., Kingsley, R. A., Connor, T. R., Harris, S. R., Parry, C. M., Al-Mashhadani, M. N., Kariuki, S., Msefula, C. L., Gordon, M. A., de Pinna, E., Wain, J., Heyderman, R. S., Obaro, S., Alonso, P. L., Mandomando, I., MacLennan, C. A., Tapia, M. D., Levine, M. M., Tennant, S. M., … Dougan, G. (2012). Intracontinental spread of human invasive Salmonella Typhimurium pathovariants in sub-Saharan Africa. *Nature Genetics*, *44*(11), 1215–1221. https://doi.org/10.1038/ng.2423

53. Pilar, A. V. C., Reid-Yu, S. A., Cooper, C. A., Mulder, D. T., & Coombes, B. K. (2012). GogB Is an Anti-Inflammatory Effector that Limits Tissue Damage during Salmonella Infection through Interaction with Human FBXO22 and Skp1. *PLoS Pathogens*, *8*(6), e1002773. https://doi.org/10.1371/journal.ppat.1002773

54. Pratap, C. B., Kumar, G., Patel, S. K., Verma, A. K., Shukla, V. K., Kumar, K., & Nath, G. (2013). Targeting of putative fimbrial gene for detection of S. Typhi in typhoid fever and chronic typhoid carriers by nested PCR. *The Journal of Infection in Developing Countries*, *7*(07), 520–527. https://doi.org/10.3855/jidc.2561

55. Pulford, C. V., Perez-Sepulveda, B. M., Canals, R., Bevington, J. A., Bengtsson, R. J., Wenner, N., Rodwell, E. V., Kumwenda, B., Zhu, X., Bennett, R. J., Stenhouse, G. E., Malaka De Silva, P., Webster, H. J., Bengoechea, J. A., Dumigan, A., Tran-Dien, A., Prakash, R., Banda, H. C., Alufandika, L., … Hinton, J. C. D. (2021). Stepwise evolution of Salmonella Typhimurium ST313 causing bloodstream infection in Africa. *Nature Microbiology*, *6*(3), 327–338. https://doi.org/10.1038/s41564-020-00836-1

56. Raffatellu, M., Wilson, R. P., Chessa, D., Andrews-Polymenis, H., Tran, Q. T., Lawhon, S., Khare, S., Adams, L. G., & Bäumler, A. J. (2005). SipA, SopA, SopB, SopD, and SopE2 Contribute to *Salmonella enterica* Serotype Typhimurium Invasion of Epithelial Cells. *Infection and Immunity*, *73*(1), 146–154. https://doi.org/10.1128/IAI.73.1.146-154.2005

57. Rana, S., Maurya, S., Mohapatra, G., Singh, S., Babar, R., Chandrasekhar, H., Chamoli, G., Rathore, D., Kshetrapal, P., & Srikanth, C. V. (2021). Activation of epigenetic regulator KDM6B by *Salmonella* Typhimurium enables chronic infections. *Gut Microbes*, *13*(1), 1986665. https://doi.org/10.1080/19490976.2021.1986665

58. Rolhion, N., Furniss, R. C. D., Grabe, G., Ryan, A., Liu, M., Matthews, S. A., & Holden, D. W. (2016). Inhibition of Nuclear Transport of NF-κB p65 by the Salmonella Type III Secretion System Effector SpvD. *PLOS Pathogens*, *12*(5), e1005653. https://doi.org/10.1371/journal.ppat.1005653

59. Rowland, R. H., Ian R. (2000). Metabolic Activities of the Gut Microflora in Relation to Cancer. *Microbial Ecology in Health and Disease*, *12*(2), 179–185. https://doi.org/10.1080/089106000750060431

60. Scanu, T., Spaapen, R. M., Bakker, J. M., Pratap, C. B., Wu, L., Hofland, I., Broeks, A., Shukla, V. K., Kumar, M., Janssen, H., Song, J.-Y., Neefjes-Borst, E. A., te Riele,

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

H., Holden, D. W., Nath, G., & Neefjes, J. (2015). Salmonella Manipulation of Host Signaling Pathways Provokes Cellular Transformation Associated with Gallbladder Carcinoma. *Cell Host & Microbe*, *17*(6), 763–774. https://doi.org/10.1016/j.chom.2015.05.002

61. Sears CL, Pardoll DM. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis 2011; 203:306–11.

62. Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

63. Shaffer, E., & Hundal, R. (2014). Gallbladder cancer: Epidemiology and outcome. *Clinical Epidemiology*, 99. https://doi.org/10.2147/CLEP.S37357

64. Sharma, A., Dwary, A. D., Mohanti, B. K., Deo, S. V., Pal, S., Sreenivas, V., Raina, V., Shukla, N. K., Thulkar, S., Garg, P., & Chaudhary, S. P. (2010). Best Supportive Care Compared With Chemotherapy for Unresectable Gall Bladder Cancer: A Randomized Controlled Study. *Journal of Clinical Oncology*. https://doi.org/10.1200/JCO.2010.29.3605

65. Sharma, A., Sharma, K. L., Gupta, A., Yadav, A., & Kumar, A. (2017). Gallbladder cancer epidemiology, pathogenesis and molecular genetics: Recent update. *World Journal of Gastroenterology*, *23*(22), 3978. https://doi.org/10.3748/wjg.v23.i22.3978

66. Shukla, V. K., Singh, H., Pandey, M., Upadhyay, S. K., & Nath, G. (2000). Carcinoma of the gallbladder—Is it a sequel of typhoid? *Digestive Diseases and Sciences*, *45*(5), 900–903. https://doi.org/10.1023/a:1005564822630

67. Smyth, G. K. (2005). limma: Linear Models for Microarray Data. In R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, & S. Dudoit (Eds.), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (pp. 397–420). Springer-Verlag. https://doi.org/10.1007/0-387-29362-0_23

68. Song, C., Lv, J., Liu, Y., Chen, J. G., Ge, Z., Zhu, J., Dai, J., Du, L.-B., Yu, C., Guo, Y., Bian, Z., Yang, L., Chen, Y., Chen, Z., Liu, J., Jiang, J., Zhu, L., Zhai, X., Jiang, Y., … for the China Kadoorie Biobank Collaborative Group. (2019). Associations Between Hepatitis B Virus Infection and Risk of All Cancer Types. *JAMA Network Open*, *2*(6), e195718. https://doi.org/10.1001/jamanetworkopen.2019.5718

69. Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

70. Stanley, T. L., Ellermeier, C. D., & Slauch, J. M. (2000). Tissue-Specific Gene Expression Identifies a Gene in the Lysogenic Phage Gifsy-1 That Affects *Salmonella enterica* Serovar Typhimurium Survival in Peyer's Patches. *Journal of Bacteriology*, *182*(16), 4406–4413. https://doi.org/10.1128/JB.182.16.4406-4413.2000

71. Stinton, L. M., & Shaffer, E. A. (2012). Epidemiology of Gallbladder Disease: Cholelithiasis and Cancer. *Gut and Liver*, *6*(2), 172–187. https://doi.org/10.5009/gnl.2012.6.2.172

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

72. Thomas, R. M., Gharaibeh, R. Z., Gauthier, J., Beveridge, M., Pope, J. L., Guijarro, M. V., Yu, Q., He, Z., Ohland, C., Newsome, R., Trevino, J., Hughes, S. J., Reinhard, M., Winglee, K., Fodor, A. A., Zajac-Kaye, M., & Jobin, C. (2018). Intestinal microbiota enhances pancreatic carcinogenesis in preclinical models. *Carcinogenesis*, *39*(8), 1068–1078. https://doi.org/10.1093/carcin/bgy073

73. Tjalsma H, Boleij A, Marchesi JR, et al. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol 2012; 10:575–82.

74. Tomasetti, C., Li, L., & Vogelstein, B. (2017). Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*, *355*(6331), 1330–1334. https://doi.org/10.1126/science.aaf9011

75. Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., Gladstone, R. A., Lo, S., Beaudoin, C., Floto, R. A., Frost, S. D. W., Corander, J., Bentley, S. D., & Parkhill, J. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*, *21*(1), 180. https://doi.org/10.1186/s13059-020-02090-4

76. Unisa, S., Jagannath, P., Dhir, V., Khandelwal, C., Sarangi, L., & Roy, T. K. (2011). Population-based study to estimate prevalence and determine risk factors of gallbladder diseases in the rural Gangetic basin of North India. *HPB*, *13*(2), 117–125. https://doi.org/10.1111/j.1477-2574.2010.00255.x

77. Upadhayay, A., Pal, D., & Kumar, A. (2022). Salmonella typhi induced oncogenesis in gallbladder cancer: Co-relation and progression. *Advances in Cancer Biology - Metastasis*, *4*, 100032. https://doi.org/10.1016/j.adcanc.2022.100032

78. Verma, S., Mohapatra, G., Ahmad, S. M., Rana, S., Jain, S., Khalsa, J. K., & Srikanth, C. V. (2015). Salmonella Engages Host MicroRNAs To Modulate SUMOylation: A New Arsenal for Intracellular Survival. *Molecular and Cellular Biology*, *35*(17), 2932–2946. https://doi.org/10.1128/MCB.00397-15

79. Welton, JohnC., Marr, JohnS., & Friedman, StephenM. (1979). ASSOCIATION BETWEEN HEPATOBILIARY CANCER AND TYPHOID CARRIER STATE. *The Lancet*, *313*(8120), 791–794. https://doi.org/10.1016/S0140-6736(79)91315-1

80. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., & Wishart, D. S. (2011). PHAST: A Fast Phage Search Tool. *Nucleic Acids Research*, *39*(suppl), W347–W352. https://doi.org/10.1093/nar/gkr485

81. Zur Hausen, H. (2002). Papillomaviruses and cancer: From basic studies to clinical application. *Nature Reviews Cancer*, *2*(5), 342–350. https://doi.org/10.1038/nrc798

82. Howlader N, Noone AM, Krapcho M, et al., editors. SEER Cancer Statistics Review. Bethesda, MD: National Cancer Institute; 1975-2014. Available from:https://seer.cancer.gov/csr/1975_2014/.

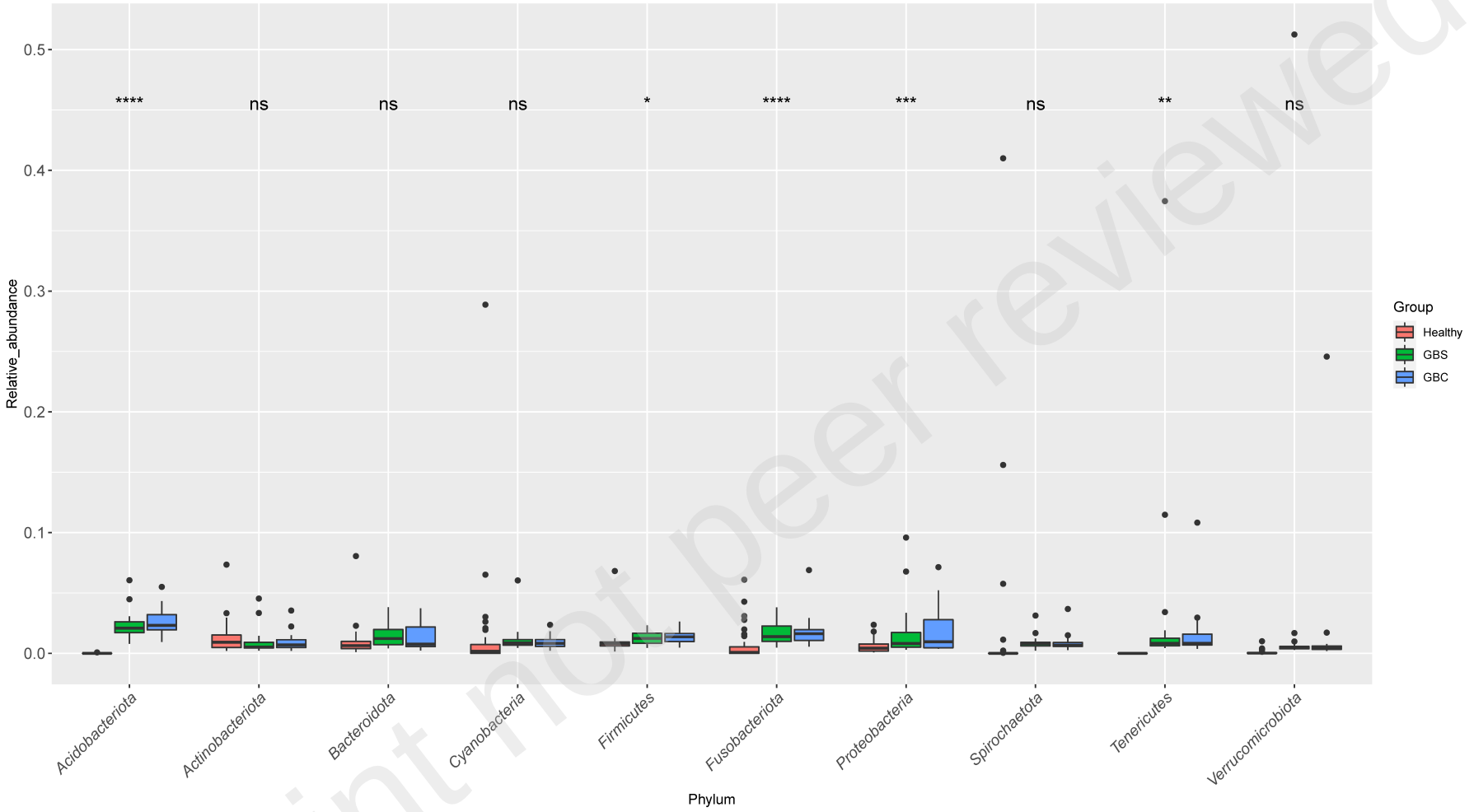
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

83. National Cancer registry programme. Consolidated report of population based cancer registries: 2012-14 [Internet]. Available online: http://ncdirindia.org/NCRP/ALL_NCRP_REPORTS/PBCR_REPORT_2012_2014/index.htm

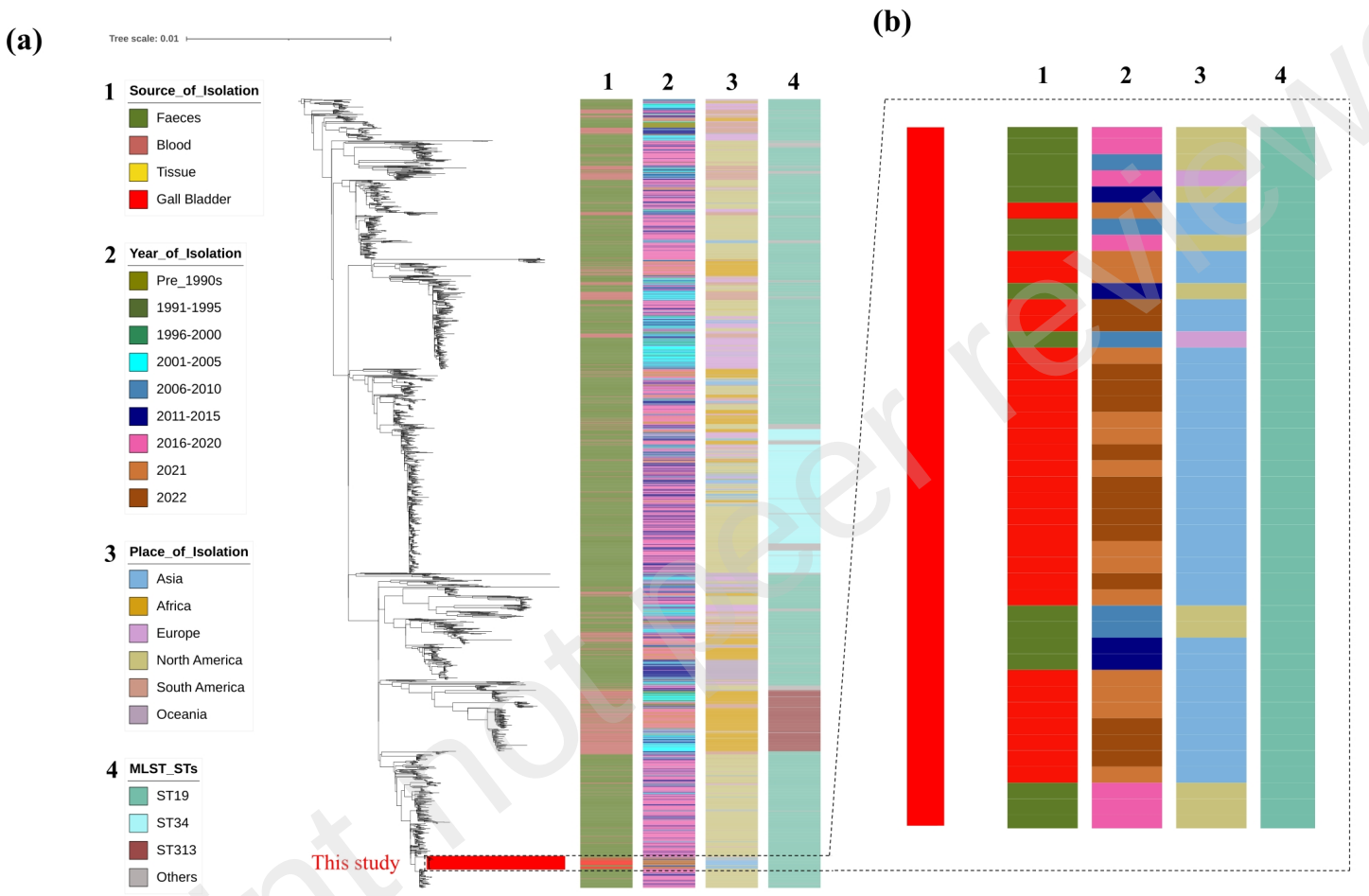
84. Yang S, Qin L, Wu P, Liu Y, Zhang Y, Mao B, Yan Y, Yan S, Tan F, Yue X, Liu H, Xue H. (2023) RNA sequencing revealed the multi-stage transcriptome transformations during the development of gallbladder cancer associated with chronic inflammation. PLoSONE, 18(3):e0283770

Preprint not peer reviewed

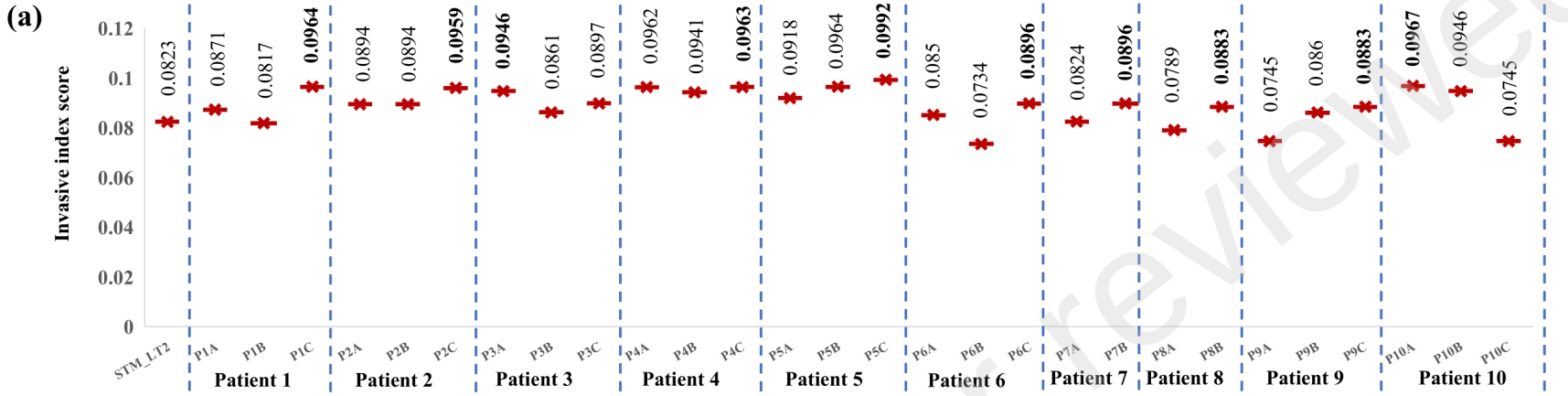
This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

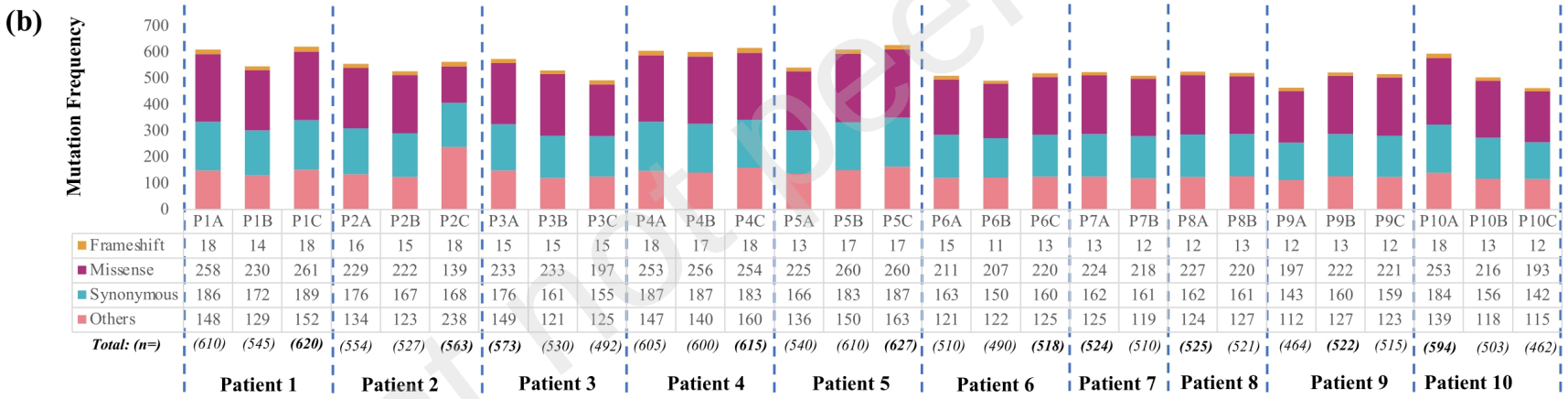


Preprint not peer reviewed



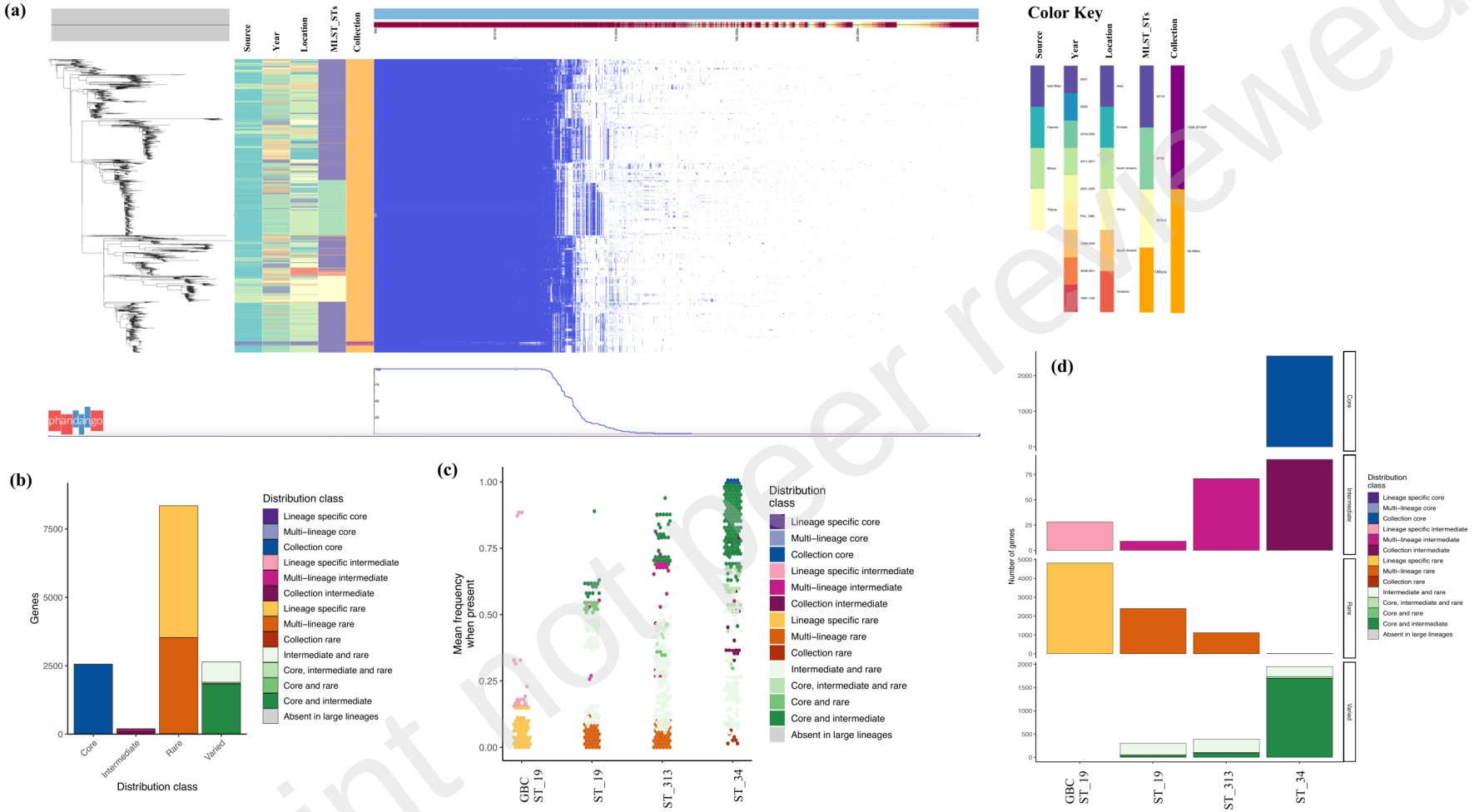
Preprint not peer reviewed



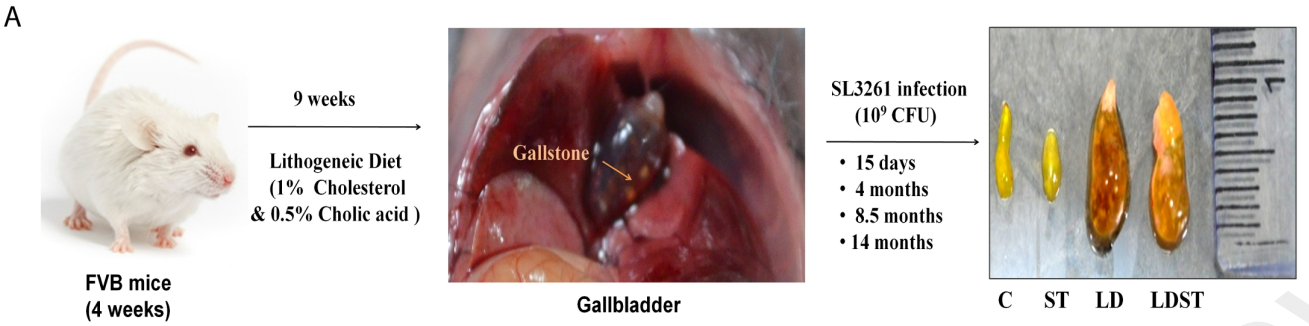


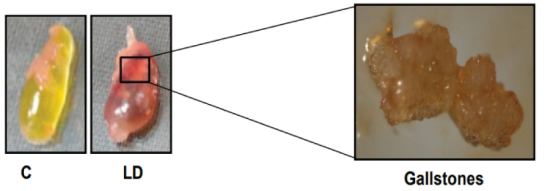
	P1A	P1B	P1C	P2A	P2B	P2C	P3A	P3B	P3C	P4A	P4B	P4C	P5A	P5B	P5C	P6A	P6B	P6C	P7A	P7B	P8A	P8B	P9A	P9B	P9C	P10A	P10B	P10C
Frameshift	18	14	18	16	15	18	15	15	15	18	17	18	13	17	17	15	11	13	13	12	12	13	12	13	12	18	13	12
Missense	258	230	261	229	222	139	233	233	197	253	256	254	225	260	260	211	207	220	224	218	227	220	197	222	221	253	216	193
Synonymous	186	172	189	176	167	168	176	161	155	187	187	183	166	183	187	163	150	160	162	161	162	161	143	160	159	184	156	142
Others	148	129	152	134	123	238	149	121	125	147	140	160	136	150	163	121	122	125	125	119	124	127	112	127	123	139	118	115
Total: (n=)	*(610)*	*(545)*	***(620)***	*(554)*	*(527)*	***(563)***	***(573)***	*(530)*	*(492)*	*(605)*	*(600)*	***(615)***	*(540)*	*(610)*	***(627)***	*(510)*	*(490)*	***(518)***	***(524)***	*(510)*	***(525)***	*(521)*	*(464)*	***(522)***	*(515)*	***(594)***	*(503)*	*(462)*

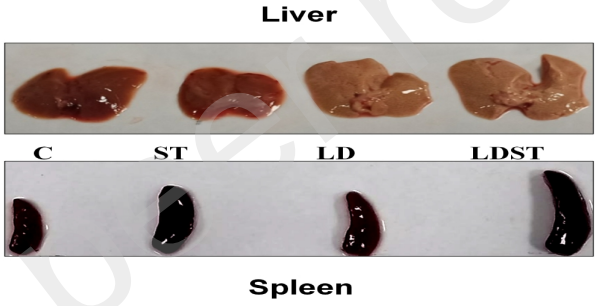
Preprint not peer reviewed

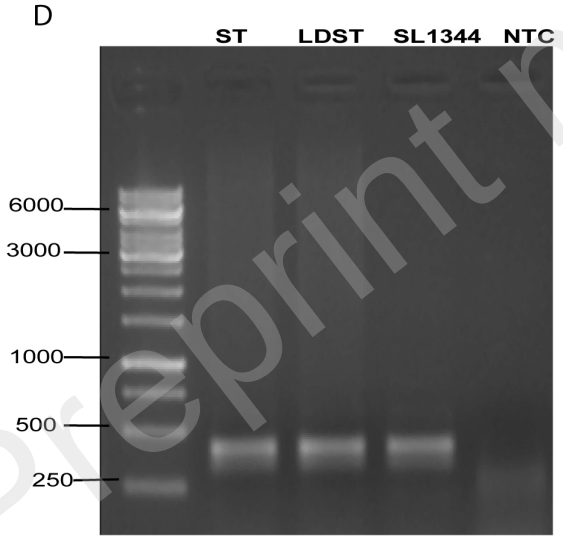


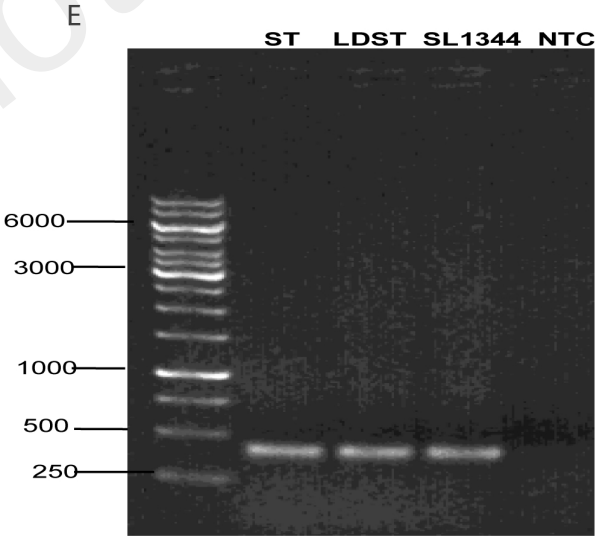
Preprint not peer reviewed





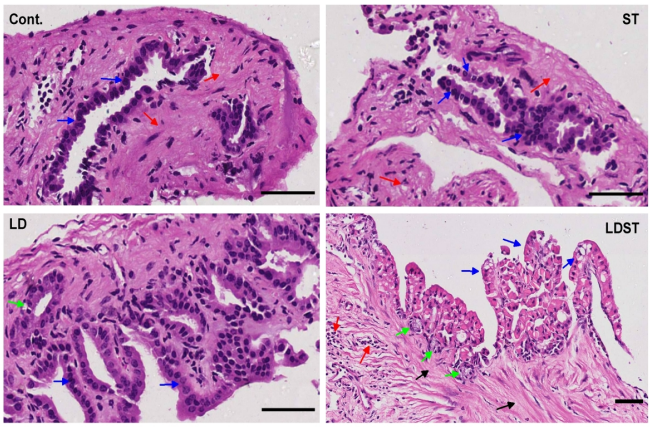


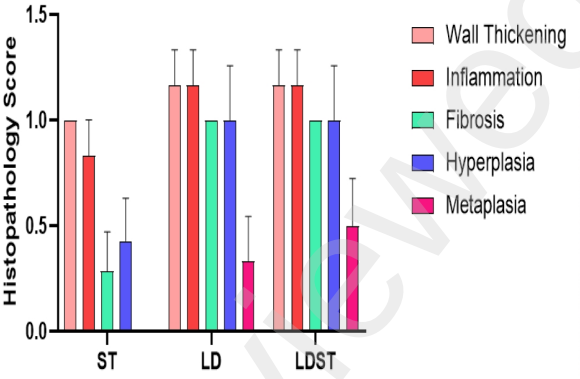




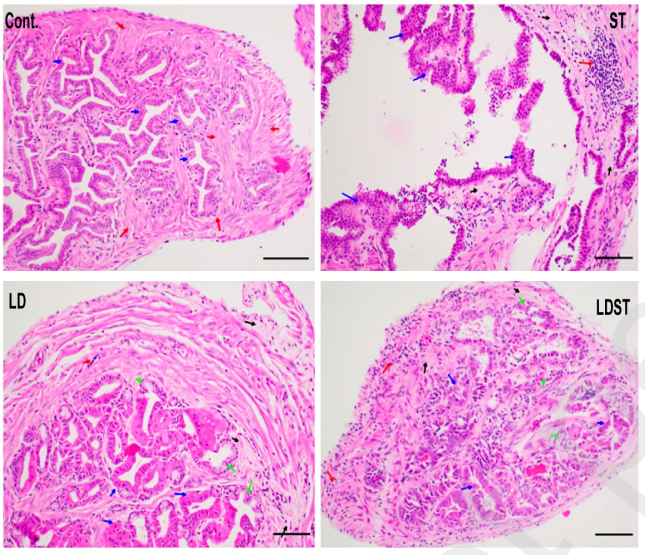
Preprint not peer reviewed

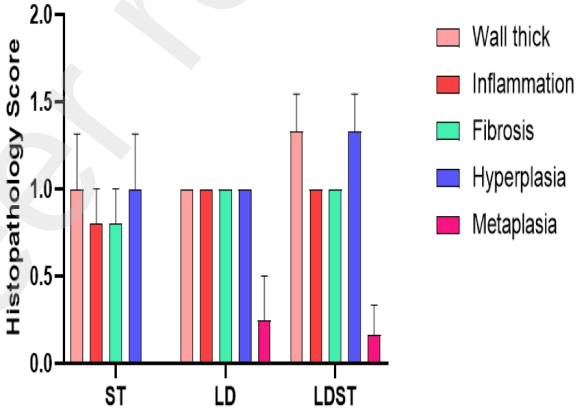
A



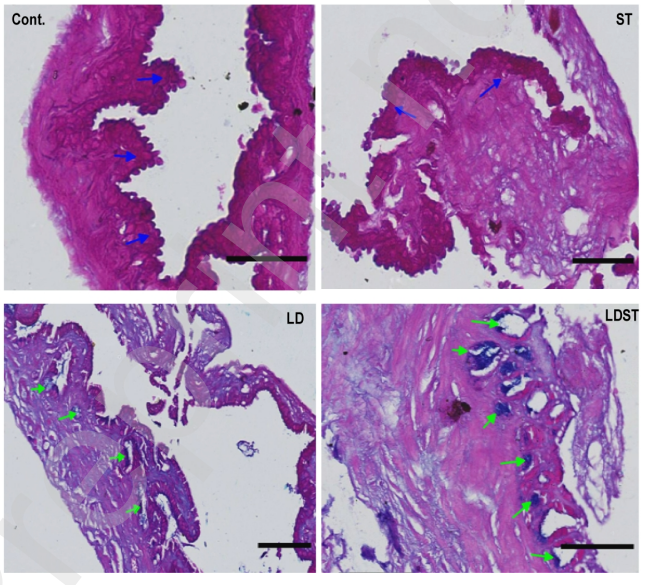


B

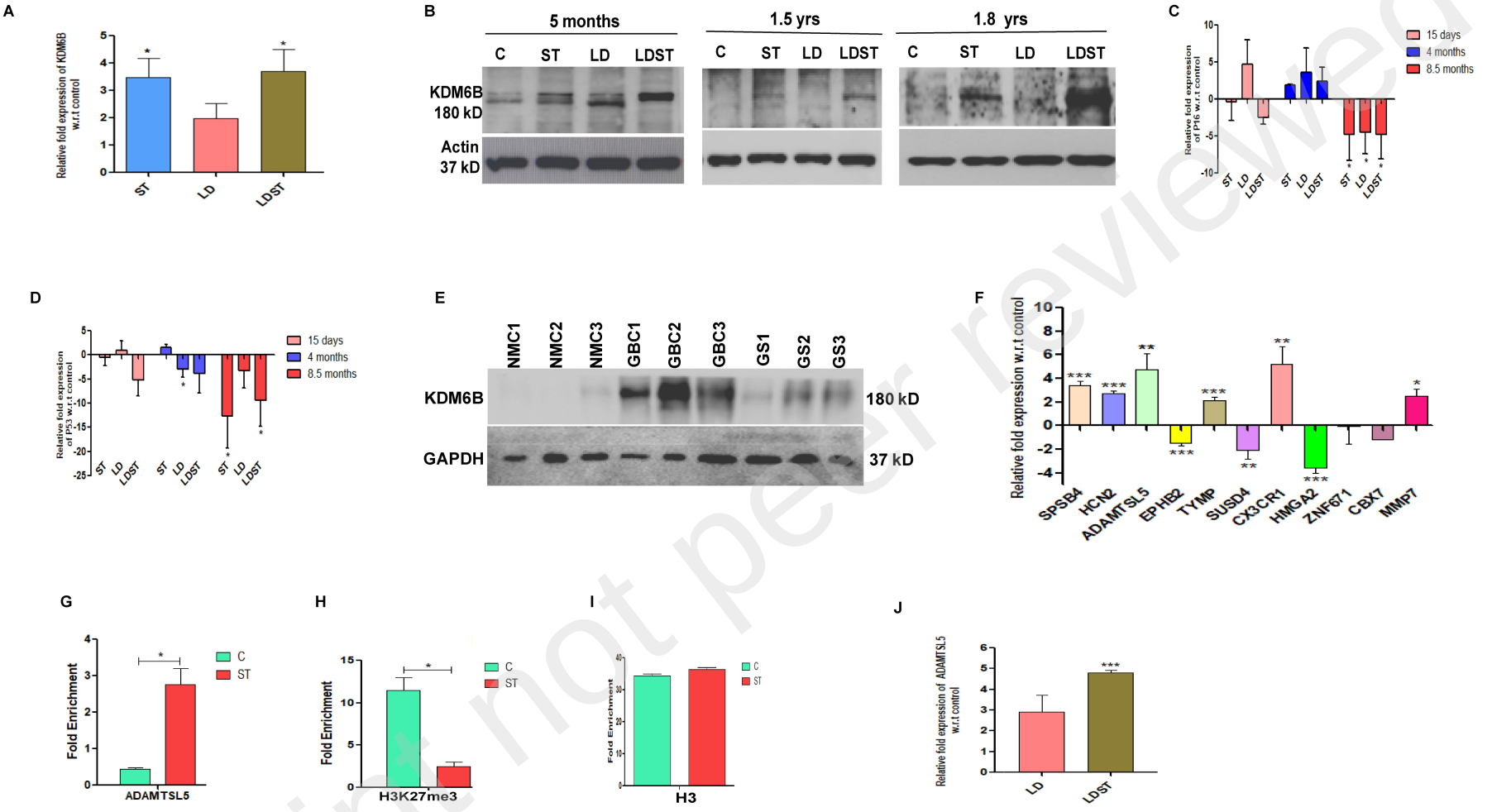




C

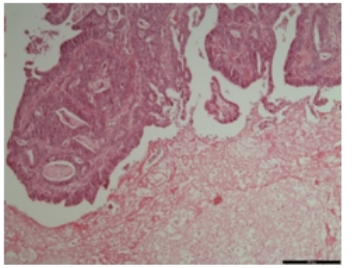


Pre-print not peer reviewed

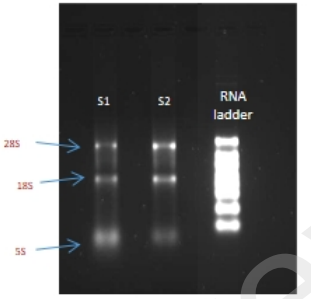


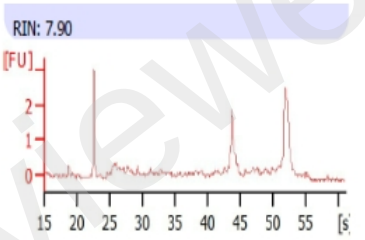
Preprint not peer reviewed

(a)

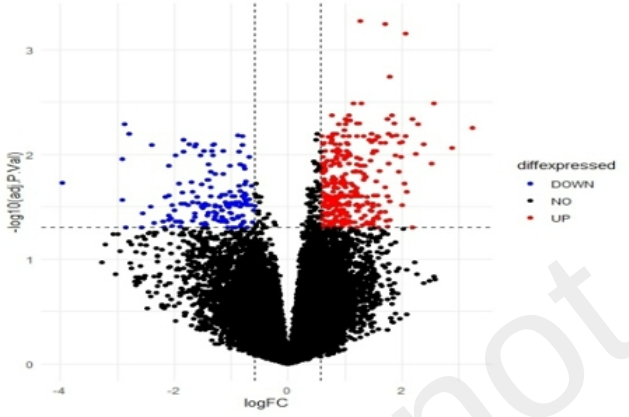


(b)

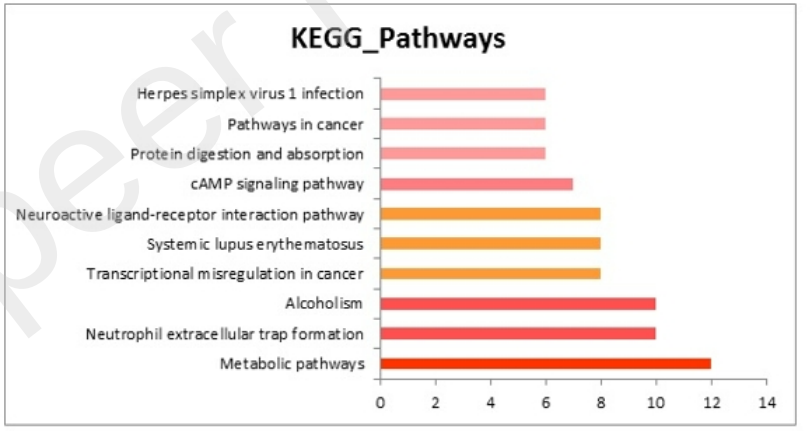




(c)



(d)



(e)

Preprint not peer reviewed

Tables

Table 1: 16S rRNA raw reads processing information of the faecal samples from GBS (n=20) and GBC (n=20) patients

Sample name	Classification	No of raw reads	No of qualified reads	Average read length
SO_9130_1	GBS (Gall bladder stone)	51766	50945	1352.6
SO_9130_11		58224	57680	1353.6
SO_9130_12		55014	53974	1196
SO_9130_16		62999	61195	1306.7
SO_9130_17		44113	38105	1220.5
SO_9130_18		35924	35580	1278.4
SO_9130_2		74288	72889	1297.2
SO_9130_21		28241	27862	1281.1
SO_9130_23		33965	33717	1350.4
SO_9130_29		41458	41077	1341.8
SO_9130_3		29574	29247	1336.3
SO_9130_31		36022	35488	1290.2
SO_9130_32		32977	32281	1264.5
SO_9130_36		55846	55453	1350.3
SO_9130_37		36446	36135	1349.3
SO_9130_39		30600	30378	1352.1
SO_9130_4		57760	57241	1327.3
SO_9130_40		22899	22719	1330.7
SO_9130_5		41918	41444	1307.5
SO_9130_8		54128	53011	1297.5
SO_9130_10	GBC (Gall bladder cancer)	57272	56042	1224.2
SO_9130_13		36056	35211	1289.1
SO_9130_14		45718	44339	1289.4
SO_9130_15		77590	70719	972.2
SO_9130_19		34073	33802	1328.8
SO_9130_20		58939	58450	1328.8
SO_9130_22		35302	35024	1348.8
SO_9130_24		24749	24525	1347.6
SO_9130_25		36856	36397	1315.5
SO_9130_26		37002	36772	1366.5
SO_9130_27		28958	28719	1366.5
SO_9130_28		29215	28909	1307.6
SO_9130_30		42628	42149	1306.6
SO_9130_33		47905	47512	1306.6
SO_9130_34		28864	28648	1306.6
SO_9130_35		50459	50052	1360.2
SO_9130_38		43271	42937	1340.4
SO_9130_6		58345	57660	1344.3
SO_9130_7		39786	39021	1344.3
SO_9130_9		55193	54430	1309.4

Table 2: Preprocessing and quality assessment of 16S rRNA raw reads of tissue sample from GBC patients (n=9)

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Patient ID	Input Read pairs	Both surviving reads	%Both Surviving Read	Forward Only Surviving Reads	%Forward Only Surviving Read	Reverse Only Surviving Reads	%Reverse Only Surviving Read	Dropped Reads	%Dropped Read	Merged reads	%merged read
P3C	276970	200064	72.23	15235	5.5	50200	18.12	11471	4.14	53228	26.6054862
P4A	291159	214779	73.77	16669	5.73	48123	16.53	11588	3.98	64811	30.17566894
P4B	278842	206969	74.22	15756	5.65	45194	16.21	10923	3.92	67637	32.6797733
P4C	290395	219426	75.56	15079	5.19	46573	16.04	9317	3.21	74824	34.09987877
P5A	261265	191532	73.31	14881	5.7	44222	16.93	10630	4.07	54024	28.20625274
P5B	307164	226657	73.79	17132	5.58	51066	16.62	12309	4.01	70844	31.25603886
P5C	273855	202276	73.86	16295	5.95	4457 8	16.28	10706	3.91	59336	29.33417707
P6A	325482	247558	76.06	17772	5.46	50613	15.55	9539	2.93	83576	33.76016933
P6B	260129	199496	76.69	14381	5.3	39018	15	7234	2.78	66430	33.29891326

Table 3: Multi locus sequence typing (MLST) allelic profile of the study isolates depicting sequence types (STs) based on seven housekeeping genes

	IDs	***ST***	***aroC***	***dnaN***	***hemD***	***hisD***	***purE***	***sucA***	***thrA***
Patient 1	P1A	19	10	7	12	9	5	9	2
	P1B	19	10	7	12	9	5	9	2
	P1C	19	10	7	12	9	5	9	2
Patient 2	P2A	19	10	7	12	9	5	9	2
	P2B	19	10	7	12	9	5	9	2
	P2C	19	10	7	12	9	5	9	2
Patient 3	P3A	19	10	7	12	9	5	9	2
	P3B	ND	ND	7	12	9	5	9	2
	P3C	19	10	7	12	9	5	9	2
Patient 4	P4A	19	10	7	12	9	5	9	2
	P4B	19	10	7	12	9	5	9	2
	P4C	19	10	7	12	9	5	9	2
Patient 5	P5A	19	10	7	12	9	5	9	2
	P5B	ND	~10	7	12	9	5	9	2
	P5C	19	10	7	12	9	5	9	2
Patient 6	P6A	19	10	7	12	9	5	9	2
	P6B	ND	10	7	12	9	ND	ND	2
	P6C	19	10	7	12	9	5	9	2
Patient 7	P7A	19	10	7	12	9	5	9	2
	P7B	19	10	7	12	9	5	9	2
Patient 8	P8A	19	10	7	12	9	5	9	2
	P8B	19	10	7	12	9	5	9	2
Patient 9	P9A	ND	10	7	12	9	ND	ND	2
	P9B	19	10	7	12	9	5	9	2

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

	P9C	19	10	7	12	9	5	9	2
Patient 10	P10A	19	10	7	12	9	5	9	2
	P10B	19	10	7	12	9	5	9	2
	P10C	ND	10	7	12	9	ND	ND	2

ND- Not determined

Table 4: Detection of plasmid and plasmid replicons in the studied genome of *S.* Typhimurium, ST- 19 isolated from GBC (n=28) with reference to *S.* Typhimurium LT2 (ST-19 serovar) as identified by PLSDB and REBASE

		plasmid P02	**plasmid pSE81-1705-3**	**plasmid 3**	**CFSAN002003 plasmid (100% identity with pSLT)**
GC content		52.6%	50.2%	52.2%	53.1%
Virulence gene		*entD*	*galU*	*entD*	*pefC, spvB, spvC , pefD, rck, pefA, pefB*
Plasmid replicons		NA	NA	NA	*IncFII, IncFI*
Patient 1	P1A	+	+	+	-
	P1B	+	+	+	-
	P1C	+	+	-	-
Patient 2	P2A	+	+	-	-
	P2B	+	+	-	-
	P2C	+	+	+	+
Patient 3	P3A	+	+	+	+
	P3B	+	+	-	-

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

	P3C	+	-	+	-
Patient 4	P4A	+	+	+	-
	P4B	+	+	-	-
	P4C	+	+	-	-
Patient 5	P5A	+	+	-	+
	P5B	+	+	-	+
	P5C	+	+	-	+
Patient 6	P6A	+	+	-	+
	P6B	+	+	-	+
	P6C	+	+	+	-
Patient 7	P7A	+	+	+	-
	P7B	+	+	+	-
Patient 8	P8A	+	+	+	-
	P8B	+	+	+	+

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

Patient 9	P9A	+	+	+	+
	P9B	+	+	+	+
	P9C	+	+	+	+
Patient 10	P10A	+	+	+	-
	P10B	+	+	+	-
	P10C	+	+	+	-

+ indicates present, - indicates absent

Table 5: Detection of prophage and prophage remnants in the studied genome of *S.* Typhimurium, ST- 19 isolated from GBC (n=28) with reference to invasive D23580 African isolate (ST-313)

	Gifsy-1	**Gifsy-2**	**Salmon_118970_Sal3**	**Def-1**	**Def-2**	**Def-3**	**Def-4**	**ST64B**
D23580 (ST313)	+	+	Not determined	+	+	+	+	+
P1A	+	+	*	+	+	+	+	+
P1B	+	+	*	+	+	+	+	+
P1C	+	+	*	+	+	+	+	+
P2A	+	+	*	+	+	+	+	+
P2B	+	*	*	+	+	+	-	+
P2C	+	*	*	+	+	+	+	+
P3A	+	+	*	+	+	+	+	+
P3B	+	+	*	+	+	*	-	+
P3C	+	+	*	+	+	*	+	+
P4A	+	+	*	+	+	+	+	+
P4B	+	+	*	+	+	+	+	+
P4C	+	+	*	+	+	+	+	+
P5A	+	*	*	+	+	+	+	+
P5B	+	+	*	+	+	+	+	+
P5C	+	+	*	+	+	+	-	+
P6A	+	*	*	+	+	+	-	+
P6B	+	+	*	*	+	+	+	+

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291

P7A	+	+	*	+	+	+	+	+
P7B	+	*	*	+	+	+	+	+
P8A	+	*	*	+	+	+	+	+
P8B	+	*	*	+	+	+	+	+
P9A	+	+	*	*	+	+	+	+
P9B	+	*	*	+	+	+	+	+
P9C	+	*	*	+	+	+	+	+
P10A	+	+	*	+	+	+	+	+
P10B	+	+	*	+	+	+	+	+
P10C	+	+	*	*	+	+	+	+
P6C	+	+	*	+	+	+	+	+

S. Typhimurium D23580- ST313 (invasive African strain), + = present, - = absent, * = Query coverage below< 35%. Gifsy-1, Gifsy-2 and **Salmon_118970_Sal3 phages were detected by PHASTER**

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=4540291